

modulate or mediate RXR function, anti-RXR antibodies can be used to determine RXR subcellular distribution patterns

claim 2: Page 47-48: 90pp: English.

Mouse retinoid X receptor (RXR) interacting protein R1P15 (R99736) is a previously unidentified orphan member of the nuclear receptor superfamily which may be involved in the complex retinoid response. It was identified using an in vivo interaction trap system for the isolation of proteins that physically interact with RXRs, esp. with the ligand binding domain of human RXR alpha. Recombinant R1P15 can be obtained using a cDNA clone (T31929) obtained from a mouse liver library. R1P15 (See also R99735 and R99737-39) can be used to modulate or mediate RXR function, and may be used therapeutically or to raise antibodies.

Sequence 446 AA:

Query Match: 100.0%; Score 2412; DB 17; Length 446;
Best Local Similarity: 100.0%; Pred. No. 7, 76-199;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MSPTSLDTVPVNGSPGSPSTATSPTEKEPDETEPPGSSSSAVIVIIPEPDP 60
1 mspstsltdtvpvngspgststspsttekepdeetppgssssaviviipepdp 60
61 EKKKKKPAFMLCHELCPVCGDKASFEHNVLSCEGCKGFFPSVHGGAEVYARQSG 120
61 ekkkkkpaFMLCHELCPVCGDKASFEHNVLSCEGCKGFFPSVHGGAEVYARQSG 120
121 TCGMFAHPPGPGVCPVCKNKAAMPPGVSTFVYVYVYVYVYVYVYVYVYVYVYV 180
121 tCGMFAHPPGPGVCPVCKNKAAMPPGVSTFVYVYVYVYVYVYVYVYVYVYVYV 180
181 GPPAASPTSSASSSGSGFEGTGTAAQPIIMIQVAAQIQNKSPSNOQPVTPWPG 240
181 gppAASPTSSASSSGSGFEGTGTAAQPIIMIQVAAQIQNKSPSNOQPVTPWPG 240
241 ADPSPTSSASSSGSGFEGTGTAAQPIIMIQVAAQIQNKSPSNOQPVTPWPG 300
241 adpsptssasssgsgfegtgtAAQPIIMIQVAAQIQNKSPSNOQPVTPWPG 300
301 TARKYNETECITELKQPTYSKQFHRAGI QVPEINPFFSPAMPEGIITAFVAILIA 360
301 tArKYNETECITELKQPTYSKQFHRAGI QVPEINPFFSPAMPEGIITAFVAILIA 360
361 INTSADPFNVGERSVVALGQVYVVALSTFKIKQVQDLPEPPMIMKIVSLIISVH 420
361 intSADPFNVGERSVVALGQVYVVALSTFKIKQVQDLPEPPMIMKIVSLIISVH 420
421 SEQVFALELQKRLPPLISEIWDVHE 446
421 seqvfaLeLqKRLPPLISEIWDVHE 446

RESULT 2
K04169
K04169 standard: Protein: 446 AA.

K04169:

14-SEP-1996 (first entry)
OR-1 orphan receptor (nuclear recep or family) protein sequence.
retinoid acid; RXR hormone response element.

K04169: rat RXR.
K04169: 05.000 A2.
K04169: 05.000 A2.

PT

92-FRR-1996

16-AUG-1995: 95W0-EP03247

16-AUG-1994: 94GB-0016536

(KARO) KARO R10 AH.

Enmark E, Gustafsson JA;

WPI: 1996-136566/14.

R PSH: 713229.

Nuclear receptor OR-1 and DNA encoding it - is involved in

modulation of retinoid acid signalling

claim 7: Fig.1: 23pp: English.

OR-1, its complex and ligands are useful in medicine, especially

to modulate retinoid acid signalling in vivo and in vitro. OR-1

binds to retinoid X receptor (RXR) and recognizes the thyroid

hormone response element DR4. Neither of OR-1 and RXR alone can

bind to DR4 but together they form a specific complex with it,

and are able to impart responsiveness to 9-cis retinoic acid to a

DR4-containing promoter. OR-1 has a similar P-box configuration

to the retinoic acid receptor, the vitamin-D receptor and the

thyroid hormone receptor. OR-1 has a mol.wt. of 50,000. The

N-terminal domain of OR-1 is 77 AA long and to a large extent

comprises a "PEST" sequence, rich in Pro, Glu, Ser, Thr and Asp

residues. The DNA binding domain consists of 68 AA including the

9 invariant cysteines characteristic of the members of the nuclear

receptor superfamily.

Sequence 446 AA:

Query Match: 98.8%; Score 2284; DB 17; Length 446;
Best Local Similarity: 98.8%; Pred. No. 2, 50-196;
Matches 438; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

1 MSPTSLDTVPVNGSPGSPSTATSPTEKEPDETEPPGSSSSAVIVIIPEPDP 60
1 mspstsltdtvpvngspgststspsttekepdeetppgssssaviviipepdp 60
61 EKKKKKPAFMLCHELCPVCGDKASFEHNVLSCEGCKGFFPSVHGGAEVYARQSG 120
61 ekkkkkpaFMLCHELCPVCGDKASFEHNVLSCEGCKGFFPSVHGGAEVYARQSG 120
121 TCGMFAHPPGPGVCPVCKNKAAMPPGVSTFVYVYVYVYVYVYVYVYVYVYVYV 180
121 tCGMFAHPPGPGVCPVCKNKAAMPPGVSTFVYVYVYVYVYVYVYVYVYVYVYV 180
181 GPPAASPTSSASSSGSGFEGTGTAAQPIIMIQVAAQIQNKSPSNOQPVTPWPG 240
181 gppAASPTSSASSSGSGFEGTGTAAQPIIMIQVAAQIQNKSPSNOQPVTPWPG 240
241 ADPSPTSSASSSGSGFEGTGTAAQPIIMIQVAAQIQNKSPSNOQPVTPWPG 300
241 adpsptssasssgsgfegtgtAAQPIIMIQVAAQIQNKSPSNOQPVTPWPG 300
301 TARKYNETECITELKQPTYSKQFHRAGI QVPEINPFFSPAMPEGIITAFVAILIA 360
301 tArKYNETECITELKQPTYSKQFHRAGI QVPEINPFFSPAMPEGIITAFVAILIA 360
361 INTSADPFNVGERSVVALGQVYVVALSTFKIKQVQDLPEPPMIMKIVSLIISVH 420
361 intSADPFNVGERSVVALGQVYVVALSTFKIKQVQDLPEPPMIMKIVSLIISVH 420
421 SEQVFALELQKRLPPLISEIWDVHE 446
421 seqvfaLeLqKRLPPLISEIWDVHE 446

interferon- α
101-
Chulpaas

RESULT 3
R74739 standard, protocol 443 AA
XX
AC R74739
XX
DT 21-JAN-1996 (first entry)
XX
DE Rat ubiquitous nuclear receptor protein.
XX
KW Ubiquitous nuclear receptor; drug design; disease diagnosis; therapy.
XX
OS Rattus rattus.
XX
PN WC9513473-A1.
XX
PD 18-MAY-1995
XX
PF 08-NOV-1994, 94W0-US12883.
XX
PR 10-NOV-1993, 9308-0152003.
XX
PA (ARCH) ARCH DEV CORP.
XX
PI Liao S, Song C;
XX
DR WPI: 1995-194092/25.
XX
DR N-PSDB: Q88761.
XX
PT New ubiquitous nuclear receptor - used to develop prods. for use in diagnostic, drug design and therapeutic applications
XX
PS Claim 5; page 144; 195pp; English.
XX
CC Isolated from rat vagina interacts with the response elements and network of regulatory elements in the promoter region of a rat receptor subfamily. The products can be used in detection, diagnosis and in screening assays for substances which interact with UNRS for use in diagnosis, drug design and therapeutic applications.
XX
SO Sequence 443 AA.

Query Match 97.2% Score 2247.5; DB 18; Length 443;
Host local similarity 97.1% Pred. No. 4.6e-193
Matches 433; Conservative 6; Mismatches 4; Indels 3; Gaps 1
XX
UY 1 MSPTSLDIPVGNSSQSPGSPSTSAISPTIFKFAETDPPGSSSSAYIVVLEPDEP 60
DB 1 MSPTSLDIPVGNSSQSPGSPSTSAISPTIFKFAETDPPGSSSSAYIVVLEPDEP 60
XX
UY 61 EFFEKATMEMLTETFAVDEKAGTETWLSLAKKALANKVAGLAAVAGAAAG 129
DB 58 EKTRPQPPHMLQPPHMLQPPHMLQPPHMLQPPHMLQPPHMLQPPHMLQPPHML 129
XX
UY 121 TGGGAAPEEYVGFPEPTFAKMTGVVSTFVSEVETGQGGGPGPSSPAANS 180
DB 116 TGGGAAPEEYVGFPEPTFAKMTGVVSTFVSEVETGQGGGPGPSSPAANS 180
XX
UY 116 TGGGAAPEEYVGFPEPTFAKMTGVVSTFVSEVETGQGGGPGPSSPAANS 177
DB 116 TGGGAAPEEYVGFPEPTFAKMTGVVSTFVSEVETGQGGGPGPSSPAANS 177
XX
UY 181 GPAAASGRTSPASSQSGSPGPGTAAQPMTCQIVAAQ GQGRSPSPDPKVPWPIG 240
DB 178 GPAAASGRTSPASSQSGSPGPGTAAQPMTCQIVAAQ GQGRSPSPDPKVPWPIG 240
XX
UY 241 ADPSSIAAGQGFAPHTFATTSVETVPAQVAPVQVQGFQPAITKASTETMLQ 300
DB 238 ADPSSIAAGQGFAPHTFATTSVETVPAQVAPVQVQGFQPAITKASTETMLQ 300
XX
UY 401 TARKVHEHETITKDFVYSKQDHRPATGVPTNIFPERSPAKRIQIDAAVALLA 360
DB 209 TARKVHEHETITKDFVYSKQDHRPATGVPTNIFPERSPAKRIQIDAAVALLA 360
XX
UY 178 GPAAASGRTSPASSQSGSPGPGTAAQPMTCQIVAAQ GQGRSPSPDPKVPWPIG 240
DB 178 GPAAASGRTSPASSQSGSPGPGTAAQPMTCQIVAAQ GQGRSPSPDPKVPWPIG 240

UY 461 INFSAPEPVEPSPVATQVVEVALVETFKPPQVQVLEPDEPMEVSLRTSSVH 420
DB 458 INFSAPEPVEPSPVATQVVEVALVETFKPPQVQVLEPDEPMEVSLRTSSVH 420
XX
UY 421 SEGVAKLEGEKELPPLSEMDVHE 446
DB 419 SEGVAKLEGEKELPPLSEMDVHE 446
XX

RESULT 4
W25035
ID W25035 standard, protocol 443 AA.
XX
AC W25035
XX
DT 15-OCT-1997 (first entry)
XX
DE Rat ubiquitous nuclear receptor polypeptide.
XX
KW Ubiquitous nuclear receptor polypeptide; DNA binding domain; therapy; drug design; diagnosis.
XX
OS Rattus rattus.
XX
PN US5639616-A.
XX
PD 17-JUN-1997.
XX
PF 10-NOV-1993, 94US-0152003
XX
PR 18-NOV-1994, 94US-042411.
XX
PR 10-NOV-1993, 9405-0152004.
XX
PA (ARCH) ARCH DEV CORP.
XX
PI Liao S, Song C;
XX
DR WPI: 1995-194092/25.
XX
DR N-PSDB: T79635.
XX
PT DNA encoding ubiquitous nuclear receptor polypeptide(s) - for producing recombinant polypeptide(s) and detecting RNA
XX
PS Claim 1; Column 77-80; 67pp; English.
XX
CC W25035 shows the amino acid sequence of a rat ubiquitous nuclear receptor polypeptide (UR). The UR is useful in assays destined to detect substances which interact with the UR which may potentially be of use therapeutically. UR is also used in diagnosis and drug design.
XX
SO Sequence 443 AA.

Query Match 97.2% Score 2247.5; DB 18; Length 443;
Host local similarity 97.1% Pred. No. 4.6e-193
Matches 433; Conservative 6; Mismatches 4; Indels 3; Gaps 1
XX
UY 1 MSPTSLDIPVGNSSQSPGSPSTSAISPTIFKFAETDPPGSSSSAYIVVLEPDEP 60
DB 1 MSPTSLDIPVGNSSQSPGSPSTSAISPTIFKFAETDPPGSSSSAYIVVLEPDEP 60
XX
UY 61 EFFEKATMEMLTETFAVDEKAGTETWLSLAKKALANKVAGLAAVAGAAAG 129
DB 58 EKTRPQPPHMLQPPHMLQPPHMLQPPHMLQPPHMLQPPHMLQPPHMLQPPHML 129
XX
UY 121 TGGGAAPEEYVGFPEPTFAKMTGVVSTFVSEVETGQGGGPGPSSPAANS 180
DB 116 TGGGAAPEEYVGFPEPTFAKMTGVVSTFVSEVETGQGGGPGPSSPAANS 180
XX
UY 116 TGGGAAPEEYVGFPEPTFAKMTGVVSTFVSEVETGQGGGPGPSSPAANS 177
DB 116 TGGGAAPEEYVGFPEPTFAKMTGVVSTFVSEVETGQGGGPGPSSPAANS 177
XX
UY 181 GPAAASGRTSPASSQSGSPGPGTAAQPMTCQIVAAQ GQGRSPSPDPKVPWPIG 240
DB 178 GPAAASGRTSPASSQSGSPGPGTAAQPMTCQIVAAQ GQGRSPSPDPKVPWPIG 240

XX Mouse CNRER-1.
 DE
 XX
 XX CNRER-1: CNRER binding factor; transcription factor; mouse; renin
 KM negative regulatory element; cardiovascular disease; hypertension;
 KM congestive heart failure; diagnosis; therapy.
 XX
 XX Mus musculus.
 OS
 XX W0955343-A1.
 PN
 XX
 XX 04-NOV-1999.
 PU
 XX
 XX 23 APR 1999. 9906-US98502.
 PF
 XX
 XX 24-APR-1998. 9805-US98297.
 PR
 XX
 XX (HGHM) BRIGHAM & WOMENS HOSPITAL, INC.
 PA
 XX
 XX Chen YE, Hortlechi M, Dzan VJ, Tamura K.
 PI
 XX
 XX WPI: 2000-062002705.
 DR
 XX N FSEB. 235040.
 XX
 XX A new regulatory element binding factor used to treat an adverse
 P1 cardiovascular condition.
 P2
 XX
 XX Example: Page 74-75: English.
 PS
 XX
 XX This sequence represents novel mouse negative regulatory element
 CC binding transcription factor CNRER-1, the sequence being deduced
 CC from an isolated cDNA clone (see 235040). The protein shows 97%
 CC identity to rat RBP-1. CNRER-1 binds to the 5' flanking region of
 CC the renin gene and modulates expression of the renin gene. A
 CC claimed method for decreasing renin expression in a mammalian cell
 CC involves contacting the cell with a CNRER-1 inhibitor (in vitro or
 CC in vivo), the inhibitor preferably being an antisense nucleic acid
 CC or a dominant negative nucleic acid of a CNRER-1 nucleic acid (see
 CC 235040-43). The CNRER-1 inhibitor can be used to decrease CNRER-1
 CC activity in a subject, especially for treating an adverse
 CC cardiovascular condition, hypertension, or congestive heart failure
 CC (claimed). Also claimed is a method for increasing heart failure
 CC activity in a mammalian cell or subject by use of a CNRER-1 nucleic
 CC acid or polypeptide, and a method for determining the level of
 CC CNRER-1 expression, an increase in expression being indicative of
 CC an increased susceptibility of a subject for developing a
 CC renin-angiotensin system mediated disorder.
 CC
 XX
 XX Sequence 445 AA:

Query Match 59.0%; Score 1387.5; DF 21; Length 445;
 Best Local Similarity 69.7%; Pred. No. 4,9e-116;
 Matches 231, Conservation 49, Mismatches 19, Indels 27, Gaps 1.
 DY 55 PPRHPRKRRKAAVAKMGGHLLVWGRKASRPHNVSSRGRKPRSVVHGAAGY 134
 DB 73 epelrphbtkhlykylhual scvghvashljyulvcpvshklttrstst 117 129
 QY 115 ACVSTSTCYAMAFPRKRGVQVGVKPKVKAAMPKCVI SPFQKPRPQKQVQVQVNS 178
 DB 131 VVSGYQHPGPHYGYRTH KPVVILKQVPMQVGVVSGVQVTKIKIKKPKVQVQVNS 199
 QY 174 EPPAASSGRPAASRPSAASGSSGSPSTVGIITAAQPMWGGV VAAQVGNRPSTVQPK 233
 DB 191 SPRVSPSPFVYVPI-----QISFQDIIAMOKIYAAQYQVQVQVTRSDRI 232
 QY 234 VTRHPRCAHQSPVPAWAAVAVAFVAVIAISVQVIVPAK VAAQVGVQVQVQVQVNS 299
 DB 233 VTRHPRCAHQSPVPAWAAVAVAFVAVIAISVQVIVPAK VAAQVGVQVQVQVQVNS 299
 QY 294 IITMTIGTAPRYNDFETGTEPRKLVYSRQVPAVLAVER INRILKSTAMPPIILHVA 358
 DB 293 IITMTIGTAPRYNDFETGTEPRKLVYSRQVPAVLAVER INRILKSTAMPPIILHVA 358

DB 293 Iecmbleterrygqasistlthdlygvdh dphlptvltg lltscvscvqlndh 352
 QY 354 EYALLIATINISAPPNVGESEVFAIQVYVPAI SYIKRKRQVQD RFFAMIMKIVSL 413
 DB 352 EYALLIATINISAPPNVGESEVFAIQVYVPAI SYIKRKRQVQD RFFAMIMKIVSL 413
 QY 414 RYSSVSHSEQVPAKLOKRPPLSLFWVHF 446
 DB 413 RYSSVSHSEQVPAKLOKRPPLSLFWVHF 445

RESULT 14

R33744 standard, Protein: 440 AA.

R33744:

23-JUL-1993 (first entry)

XR2.

5E
 KM hraf alpha, thyroid receptor beta, hraf beta, diuretic thyroid receptor;
 KM hraf, retinoid X receptor-alpha; hraf, alpha, growth factor; ligand;
 KM transcription-activation; response element.

Homo sapiens.

W09306215-A.

01 APR 1993.

08-SEP 1992: 92W0-US07570.

17-SEP-1991: 91US-0761068.

(SAIK) SAIK INST BIOLOGICAL SCIENCES.

Berangeror UK, Evans PM, Giquerre V, Mangelsdorf DJ;

Ong ES, Oro AE, Yao TP;

WPI: 1993-117536/14.

N-PSDB: Q39088.

DNA encoding proteins of thyroid-stimulating hormone receptor

super-family, useful for screening for agonists/antagonists of

such receptors

PS

XX

XX

XX

XX

XX

XX

XX

XX

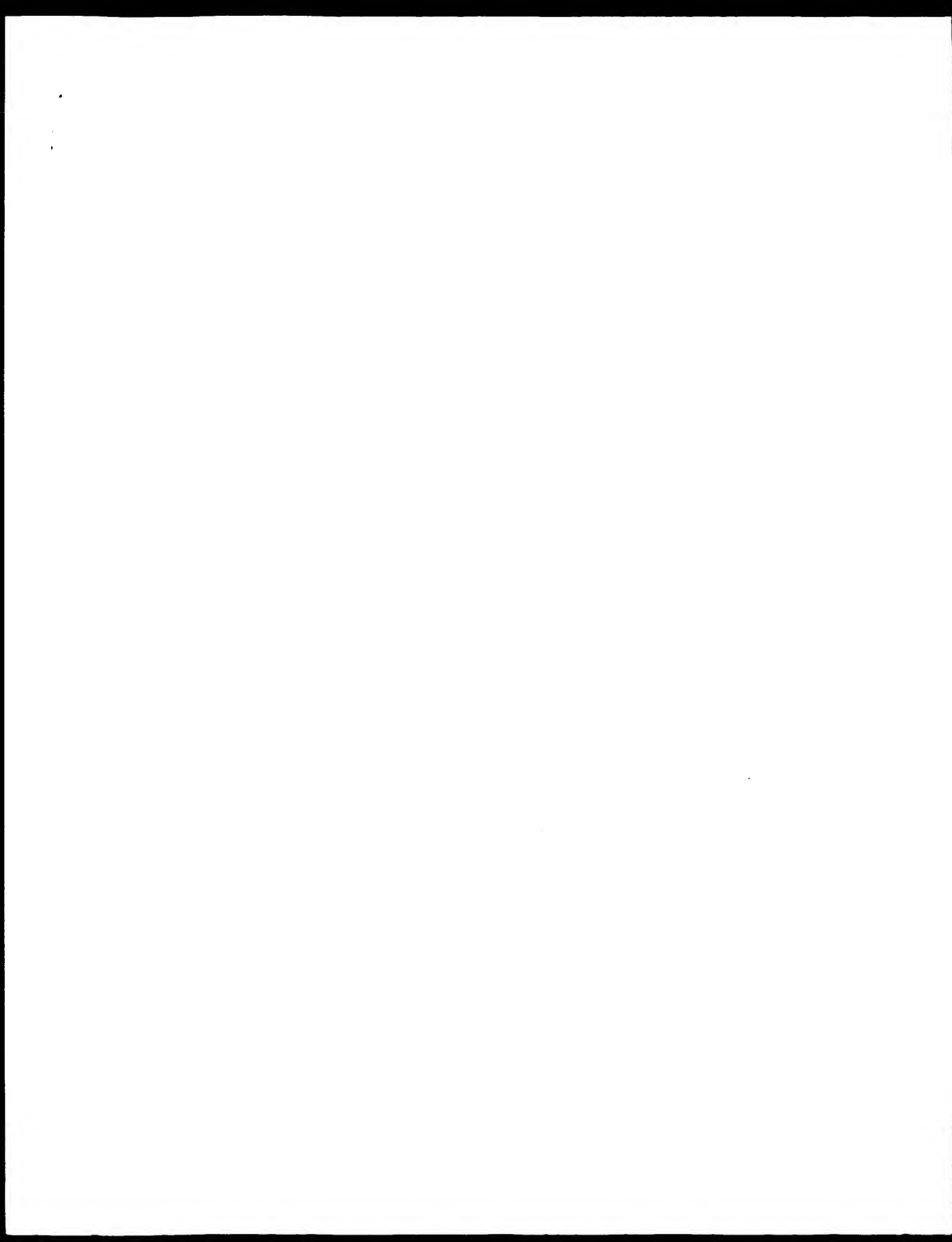
XX

XX

XX

XX

Query Match 59.9%; Score 1387.5; DF 14; Length 440;
 Best Local Similarity 62.9%; Pred. No. 8.1e-116;
 Matches 221, Conservation 48, Mismatches 94, Indels 24, Gaps 5;
 QY 4 PISLITIVVNSQGSTASPTIKKQVETDPPQSSSSAYIVVILPEDEP--- 50
 DB 4 PISLITIVVNSQGSTASPTIKKQVETDPPQSSSSAYIVVILPEDEP--- 50



GenInfo version 4.5
Copyright (c) 1998 - 2000 CompuGen Ltd.

OM protein - protein search using SW model

Run on: April 25, 2001, 09:55:08, Search time 13.35 seconds
(without alignments)
614,167 Million cell updates/sec

Title: US-09-365-576-3

Perfect score: 2312

Sequence: 1 MSSTSSSTTFVWVSPVPPV

Scoring table:

BL/BL0062
Gapop 10.0, Gapox 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 6

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents, AA:

- 1: 100-0-172-652-3
- 2: 100-0-172-652-3
- 3: 100-0-172-652-3
- 4: 100-0-172-652-3
- 5: 100-0-172-652-3
- 6: 100-0-172-652-3

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2312	100.0	446	2	US-08-372-652-3
2	2312	100.0	446	6	PCT US95/17011-2
3	2247.5	97.2	443	1	US-08-342-411A-4
4	2093.5	98.0	451	1	US-08-330-519-2
5	2093.5	98.0	451	1	US-08-330-281-2
6	2093.5	98.0	451	2	US-08-646-210-2
7	2093.5	98.0	451	5	PCT US95/17011-2
8	2093.5	98.0	451	5	PCT US95/17011-2
9	2093.5	98.0	451	5	PCT US95/17011-2
10	1955.5	92.4	447	1	US-08-312-111A-2
11	1395	60.3	440	1	US-08-333-458-8
12	1395	60.3	440	1	US-08-463-694-8
13	1395	60.3	440	1	US-08-694-501-8
14	1281.5	55.4	432	2	US-08-476-120-2
15	1281.5	55.4	432	2	PCT US94/07266-2
16	704	30.4	764	4	US-06-144-750-20
17	699.5	27.9	459	3	US-09-144-759-18
18	645.5	27.9	459	3	US-08-372-181-2
19	645.5	27.9	459	3	US-09-469-721-2
20	645.5	27.9	459	5	PCT US95/17011-2
21	641	27.6	422	1	US-08-496-611-2
22	638.5	27.6	484	2	US-08-372-652-3
23	638.5	27.6	484	5	PCT US95/16411-1
24	626.5	27.1	451	5	PCT US95/16411-2
25	626.5	27.1	451	5	PCT US95/16411-2
26	571	24.7	452	2	US-08-592-483-2
27	571	24.7	452	2	US-08-592-483-2
28	571	24.7	452	5	PCT US92/02200A-4
29	563	24.4	462	6	5171671-2
30	556.5	24.1	454	6	526432-2
31	555.5	24.0	403	2	US-08-592-484-4
32	511.5	23.1	797	2	US-08-096-729B-2
33	511.5	23.1	797	5	PCT US92/02200A-2
34	525	22.7	448	6	522406-2
35	525	22.7	448	6	522406-2
36	497.5	21.5	410	6	548145-2
37	495	21.4	455	6	522406-4
38	479	20.7	348	1	US-08-458-461-10
39	479	20.7	348	1	US-08-458-461-10
40	479	20.7	348	1	US-07-848-360C-10
41	478	19.8	453	2	US-07-848-360C-2
42	451	19.5	446	1	US-07-952-800-4
43	447	19.3	548	1	US-08-333-508-2
44	447	19.2	518	1	US-08-453-504-2
45	447	19.3	548	1	US-08-694-501-2

ALIGNMENTS

RESULT 1
US-08-372-652-3
Sequence 3: Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David
INVENTOR: Choi, Hong-Sik
TITLE OF INVENTION: METHOD AND SYSTEM FOR IDENTIFYING AND CLASSIFYING PROTEINS AND NUCLEIC ACIDS
NAMES OF INVENTORS: Choi, Hong-Sik
CORRESPONDENCE ADDRESS:
ADDRESS: 225 Franklin Street, Suite 4100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn release #1.0, Version #1.30
CHECKED APPLICATION DATA:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 36,162
REFERENCE/CI/CI/CI NUMBER: 08/024,001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8406
PRIORITY: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 446 amino acids
TYPE: amino acid
STRANDNESS: 504 relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US 08 372 652-3

Query Match: 100.0% Score 2312 DB 2: Length 446:
Best local similarity: 100.0% Prod. No. 1,10-216:
Matches 446, Conservative 0, Mismatches 0, Indels 0, Gaps 0:
1 MSSTSSSTTFVWVSPVPPV
|||||


```

: STREET: 126 East Lincoln Avenue
: CITY: Rahway
: STATE: New Jersey
: COUNTRY: US
: ZIP: 07065-0907
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: pc-pos/ms-dos
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: FILING DATE:
: APPLICATION NUMBER: US/08/330,283
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dolan, Catherine A.
: REGISTRATION NUMBER: 46,502
: REFERENCE/DCKET NUMBER: 19327
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (908) 594-4283
: TELEFAX: (908) 594-4720
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 461 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US 08/330-283-2

Query Match: 88.08; Score 2033.5; Db 1; Length 461;
Post Local Similarity: 87.28; Pred. No. 4,56-189;
Matches 402; Conservative 17; Mismatches 27; Indels 15; Gaps 5;

QY 1 MSSP-TSSLPVPGNSPGDSTSPSTKREGQ-----TDP-PSGSSSAIV-- 51
DB 1 MSSP-TSSLPVPGNSPGDSTSPSTKREGQ-----TDP-PSGSSSAIV-- 60
QY 52 VLEPDEPERKRRKGAVALMCHETVVGSGKASQAHNVLSGQKGPSSVHGA 111
DB 61 VLPDEPERKRRKGAVALMCHETVVGSGKASQAHNVLSGQKGPSSVHGA 120
QY 112 GRYAGRGSGVGMALPFRKQGLRLKTHAKMRQGVLSSEVLRREKLGQGG 167
DB 121 KRYAGRGSGVGMALPFRKQGLRLKTHAKMRQGVLSSEVLRREKLGQGG 180
QY 168 -GPPPTSTFAASSSPGPAASSTSTSEVSSQSGEPCTGTAAQPTMTGGLVAAAGVGNK 225
DB 181 USQSPVGVGSSSSASSTSPASSTSEVSSQSGEPCTGTAAQPTMTGGLVAAAGVGNK 240
QY 226 PPSDQPPVTVIWMIGALGQSPVAGQGPAPHTETATISVGLVTPAKVYVGLQCPHQ 285
DB 241 PPSDQPPVTVIWMIGALGQSPVAGQGPAPHTETATISVGLVTPAKVYVGLQCPHQ 300
QY 286 IALIKASTIEMLQTAARRVNHETCTTFLKQFTYSKIDPFRAGLVEFNPITFEESKAM 345
DB 301 IALIKASTIEMLQTAARRVNHETCTTFLKQFTYSKIDPFRAGLVEFNPITFEESKAM 360
QY 346 KRIAGDAEVALIATINIPSDRINVOEISREVALDOQVVEALLSVTRKQDQILRRR 405
DB 361 KRIAGDAEVALIATINIPSDRINVOEISREVALDOQVVEALLSVTRKQDQILRRR 420
QY 406 MLKIVSTRILSSVSHVFAIKGQNKIPPLSEIMDVE 446
DB 421 MLKIVSTRILSSVSHVFAIKGQNKIPPLSEIMDVE 461

```

```

: Patent No. 5939322
: GENERAL INFORMATION:
: APPLICANT: Friedman, Ethan
: APPLICANT: Holloway, M. Katharine
: APPLICANT: Rodan, Gideon
: APPLICANT: Rutledge, Su Jane
: APPLICANT: Schmidt, Azriel
: APPLICANT: Vogel, Robert
: TITLE OF INVENTOR: METHOD FOR FINDING PESTICIDE INITIATORS
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merck & Co., Inc.
: STREET: 126 East Lincoln Avenue
: CITY: Rahway
: STATE: New Jersey
: COUNTRY: US
: ZIP: 07065-0907
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: pc-pos/ms-dos
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: FILING DATE: 14-MAY-1996
: APPLICATION NUMBER: 08/330,283
: PRIOR APPLICATION NUMBER:
: FILING DATE: 27-OCT-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Dolan, Catherine A.
: REGISTRATION NUMBER: 46,502
: REFERENCE/DCKET NUMBER: 19327
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (908) 594-4283
: TELEFAX: (908) 594-4720
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 461 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US 08/330-283-2

Query Match: 88.08; Score 2033.5; Db 2; Length 461;
Post Local Similarity: 87.28; Pred. No. 4,56-189;
Matches 402; Conservative 17; Mismatches 27; Indels 15; Gaps 5;

QY 1 MSSP-TSSLPVPGNSPGDSTSPSTKREGQ-----TDP-PSGSSSAIV-- 51
DB 1 MSSP-TSSLPVPGNSPGDSTSPSTKREGQ-----TDP-PSGSSSAIV-- 60
QY 52 VLEPDEPERKRRKGAVALMCHETVVGSGKASQAHNVLSGQKGPSSVHGA 111
DB 61 VLPDEPERKRRKGAVALMCHETVVGSGKASQAHNVLSGQKGPSSVHGA 120
QY 112 GRYAGRGSGVGMALPFRKQGLRLKTHAKMRQGVLSSEVLRREKLGQGG 167
DB 121 KRYAGRGSGVGMALPFRKQGLRLKTHAKMRQGVLSSEVLRREKLGQGG 180
QY 168 -GPPPTSTFAASSSPGPAASSTSTSEVSSQSGEPCTGTAAQPTMTGGLVAAAGVGNK 225
DB 181 USQSPVGVGSSSSASSTSPASSTSEVSSQSGEPCTGTAAQPTMTGGLVAAAGVGNK 240
QY 226 PPSDQPPVTVIWMIGALGQSPVAGQGPAPHTETATISVGLVTPAKVYVGLQCPHQ 285
DB 241 PPSDQPPVTVIWMIGALGQSPVAGQGPAPHTETATISVGLVTPAKVYVGLQCPHQ 300
QY 286 IALIKASTIEMLQTAARRVNHETCTTFLKQFTYSKIDPFRAGLVEFNPITFEESKAM 345

```

Db 301 IALLKASTITITMTETAPRYNHECTITFLKPTYSKDEHPAGLOVEFINDEFFSRAM 360

QY 346 RPLGIDAEVALLIATINISADPNVQPPSPVPAIQQPVVATISVTPKPPQPPR 405

Db 361 RPLGIDAEVALLIATINISADPNVQPPSPVPAIQQPVVATISVTPKPPQPPR 420

QY 406 MLKRVSLPTLSSVHSEGVFALRLDQKLPILSETIMVHE 446

Db 421 MLKRVSLPTLSSVHSEGVFALRLDQKLPILSETIMVHE 461

RESULT 7

PCT-US95-13924-2

Sequence 2, Application PC/TUS9513924

GENERAL INFORMATION:

APPLICANT: Friedman, Eitan

APPLICANT: Holloway, M. Katharine

APPLICANT: Rodan, Gideon

APPLICANT: Rutledge, Su Jane

APPLICANT: Vogel, Robert

TITLE OF INVENTION: METHOD FOR FINDING RECEPTOR POTENTIATORS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.

STREET: 126 East Lincoln Avenue

CITY: Rahway

STATE: New Jersey

COUNTRY: US

ZIP: 07065-0907

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/13924

FILING DATE:

CLASSIFICATION:

AUTHOR/AGENT INFORMATION:

NAME: Quagliato, Carol S.

REGISTRATION NUMBER: 35,330

REFERENCE/WORK NUMBER: 19427 PCT

TELEPHONE: (908) 594 3809

TELEFAX: (908) 594-4720

INFORMATION FOR SEQ ID NO: 2

SEQUENCE CHARACTERISTICS:

LENGTH: 461 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

PCT-US95-13924-2

Query Match: 88.0%, Score 2033.5, DB 5, Length 461

Best Local Similarity: 87.2%, Fred No. 4.5e-189

Matches: 402, Conservative: 17, Mismatches: 27, Indels: 15, Gaps: 5.

QY 1 MSSP-TSSLUTPVQNGSPQSTASPTIKRGGT-----TDP-POSEGSSSAIV-- 51

Db 1 MSSP-TSSLUTPVQNGSPQSTASPTIKRGGT-----TDP-POSEGSSSAIV-- 60

QY 52 VITPP 111

Db 61 VITPP 120

QY 113 STVATSSSTTGMATMNRKTA THTPTCAATPQVVAATACAPPELQVQSS 167

Db 113 STVATSSSTTGMATMNRKTA THTPTCAATPQVVAATACAPPELQVQSS 167

Db 121 PVACTPGCTGMDAMPKQQGSPVPPVAVPQGVLSRPPKRTKQVQGESQS 180

QY 168 -QPPPEPMASSSGPASPQTEENSSQCECEQIQTAAQIMQIVAAQIQVNR 225

Db 181 QSSVAPDASSSSASCPASPCSPASQSGSPPEVQIAAQIMQIVAAQIQVNR 240

QY 226 EESDQFYVIMPLADPSPAPPPGFAHPTIATISVGTTPVAPQVPPHQAQGEED 285

Db 241 EESDQFYVIMPLADPSPAPPPGFAHPTIATISVGTTPVAPQVPPHQAQGEED 300

QY 286 IALLKASTITITMTETAPRYNHECTITFLKPTYSKDEHPAGLOVEFINDEFFSRAM 345

Db 301 IALLKASTITITMTETAPRYNHECTITFLKPTYSKDEHPAGLOVEFINDEFFSRAM 360

QY 346 RPLGIDAEVALLIATINISADPNVQPPSPVPAIQQPVVATISVTPKPPQPPR 405

Db 361 RPLGIDAEVALLIATINISADPNVQPPSPVPAIQQPVVATISVTPKPPQPPR 420

QY 406 MLKRVSLPTLSSVHSEGVFALRLDQKLPILSETIMVHE 446

Db 421 MLKRVSLPTLSSVHSEGVFALRLDQKLPILSETIMVHE 461

RESULT 8

PCT-US95-13931-2

Sequence 2, Application PC/TUS9513931

GENERAL INFORMATION:

APPLICANT: Friedman, Eitan

APPLICANT: Holloway, M. Katharine

APPLICANT: Rodan, Gideon

APPLICANT: Schmidt, Azriel

TITLE OF INVENTION: USE OF RECEPTOR POTENTIATORS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.

STREET: 126 East Lincoln Avenue

CITY: Rahway

STATE: New Jersey

COUNTRY: US

ZIP: 07065-0907

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/13931

FILING DATE:

CLASSIFICATION:

AUTHOR/AGENT INFORMATION:

NAME: Quagliato, Carol S.

REGISTRATION NUMBER: 35,330

REFERENCE/WORK NUMBER: 19416 PCT

TELEPHONE: (908) 594-3809

TELEFAX: (908) 594-4720

INFORMATION FOR SEQ ID NO: 2

SEQUENCE CHARACTERISTICS:

LENGTH: 461 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

PCT-US95-13931-2

Query Match: 88.0%, Score 2033.5, DB 5, Length 461

Best Local Similarity: 87.2%, Fred No. 4.5e-189

Matches: 402, Conservative: 17, Mismatches: 27, Indels: 15, Gaps: 5.

[illegible]


```

CORRESPONDENCE ADDRESS:
ADDRESSER: pretty, Schroeder, Hruschmann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: US
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM pc compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,694
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US-07/761,063
FILING DATE: 17 SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Rottier Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8943
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 545-9001
TELEFAX: (619) 545-8949
INFORMATION FOR SEQ. ID NO.: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-694-8

Query Match      60.38: Score 1395; DB 1; Length 440;
Host Local Similarity 63.18; Prod. No. 3.7e-127;
Matches 284; Conservative 54; Mismatches 84; Indels 40; Gaps 7;

QY 4 PTSSLDTPVQ-NSPOPSAISPTIKRGGDTDPGSSSS--AVIVLLEDP 60
DB 14 PDSAVEIMKPKAOJASSQAQGSCHLREFAR---MPSANATAPATLTRAPEP 69
QY 61 -----ERKKRKAIAKMLHETLFWVKASAFYVNLSCDCKGFFRSVVGAGPVAG 115
DB 70 TELPQKFRKQAFKMKGNELGKASVGHKASDFHNVLSLQDGFPRSVLQ--AHY 127
QY 117 KGSSTVGMATMPPKVTETFAFSTKFAVPPGVVSTFGLPKFPLKQVQGGPPPS 176
DB 128 HSGGHCMDIYMKRKKQCEPLKRCAGMPEEVLSEQQLIKKIKROBEQAHATSP 187
QY 177 ASSSGPRAASVGTSEASSGSGSDEHGLTFAAEFLMIOGVAAQLCKNKSFSTQNV 236
DB 188 RRS-PTQITP-----QLSPEDCMFPFVAAQVQGNRSPSSDRLPVTP 230
QY 217 WPLCADPQSRKAGQRFATHTLATISVGEIVDFAKGVPEGLQIAGFDQATIKAST 296
DB 211 PDSAVEIMKPKAOJASSQAQGSCHLREFAR---MPSANATAPATLTRAPEP 296
QY 241 WPMAPDHSRFAKQGRFAHFTLATISVGEIVDFAKQIDPGLQISREDDQALIKT 290
QY 297 MLDIARVYNEFTETTELKQFTYSKRIPIRACIQVFTNPITFPRSRARBLG 356
DB 291 MLETSKRYNQSSTIELKQFSNRPFAKQIQVFTNPITFPRSRARBLG 350
QY 457 LILAINIPSAIDQNVQDSRVEAFIQGVYVALLSYTLKQPLQGFPMIMKIVSL 416
DB 451 LILAINIPSAIDQNVQDSRVEAFIQGVYVALLSYTLKQPLQGFPMIMKIVSL 416
QY 417 SSVSRKQVAFIRIQKIPPLSLWVHE 446
DB 411 SSVSRKQVAFIRIQKIPPLSLWVHE 440

RESULT 14

```

```

US-08-694-501-8
Sequence 8; Application US/08694501
Patent No. 5710004
GENERAL INFORMATION:
APPLICANT: EVANS Ph.D., RONALD M.
APPLICANT: MANGELSDORF Ph.D., DAVID J.
APPLICANT: ONG MS., ESTELITA S.
APPLICANT: ONG Ph.D., ANTHONY E.
APPLICANT: HOKEMEYER Ph.D., TWE K.
APPLICANT: GIDDERE Ph.D., VINCENT NMN
APPLICANT: YAO Mr., TSO-PAN
TITLE OF INVENTION: NOVEL RECEPTORS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: pretty, Schroeder, Hruschmann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: US
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM pc compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,501
FILING DATE: 07-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/444,358
FILING DATE:
APPLICATION NUMBER:
APPLICATION NUMBER: US/07/761,068
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rottier Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8946
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ. ID NO.: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-694-501-8

Query Match      60.38: Score 1395; DB 1; Length 440;
Host Local Similarity 63.18; Prod. No. 3.7e-127;
Matches 284; Conservative 54; Mismatches 84; Indels 40; Gaps 7;

QY 4 PTSSLDTPVQ-NSPOPSAISPTIKRGGDTDPGSSSS--AVIVLLEDP 60
DB 14 PDSAVEIMKPKAOJASSQAQGSCHLREFAR---MPSANATAPATLTRAPEP 69
QY 61 -----ERKKRKAIAKMLHETLFWVKASAFYVNLSCDCKGFFRSVVGAGPVAG 115
DB 70 TELPQKFRKQAFKMKGNELGKASVGHKASDFHNVLSLQDGFPRSVLQ--AHY 127
QY 117 KGSSTVGMATMPPKVTETFAFSTKFAVPPGVVSTFGLPKFPLKQVQGGPPPS 176
DB 128 HSGGHCMDIYMKRKKQCEPLKRCAGMPEEVLSEQQLIKKIKROBEQAHATSP 187
QY 177 ASSSGPRAASVGTSEASSGSGSDEHGLTFAAEFLMIOGVAAQLCKNKSFSTQNV 236
DB 188 RRS-PTQITP-----QLSPEDCMFPFVAAQVQGNRSPSSDRLPVTP 230
QY 217 WPLCADPQSRKAGQRFATHTLATISVGEIVDFAKGVPEGLQIAGFDQATIKAST 296
DB 211 PDSAVEIMKPKAOJASSQAQGSCHLREFAR---MPSANATAPATLTRAPEP 296
QY 241 WPMAPDHSRFAKQGRFAHFTLATISVGEIVDFAKQIDPGLQISREDDQALIKT 290
DB 241 WPMAPDHSRFAKQGRFAHFTLATISVGEIVDFAKQIDPGLQISREDDQALIKT 290

```


QY 297 M1Q7APRYNHETPOTITIKPTYSKPDHPNACIOWEETINIFERSAMPPLCITONAYA 356
 DB 291 M1TFSRNFSSFTITLKCTSYNEEDFAAGLVEFINIFERSAMNLOINDAEFA 350
 QY 357 TITAINIFSAHPNPVPSRPVAFQDPVEALLSTPEKPGDOLPEPMKIVSLPTL 416
 DB 451 TITAINIFSAHPNPVQDQIVPRIGHTYVFAIHAYVSIHHHIDLMFEMKLKLSLPTL 410
 QY 417 SSVSHQVPAFPIQKPKLPPLSLIWDYHE 446
 DB 411 SSVSHQVPAFLRQDKKLPPLSLIWDYHE 440

RESULT 14
 US-08-466-120-2
 Sequence 2: Application US/08466120
 Patent No. 5869284
 GENERAL INFORMATION:
 APPLICANT: CAO, ET AL.
 TITLE OF INVENTION: Peptideic Acid Receptor Epsilon
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CARELLA, RYNE, BAIN, GILFILLAN,
 STREET: 6 BECKER FARM ROAD
 CITY: ROSHLAND
 STATE: NEW JERSEY
 COUNTRY: USA
 ZIP: 07068
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08466120
 FILING DATE: June 6, 1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/07266
 FILING DATE: 24 JUN 94
 ATTORNEY/AGENT INFORMATION:
 NAME: FERRARO, GREGORY D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/AGENT NUMBER: 325900-354
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 433 AMINO ACIDS
 TYPE: AMINO ACID
 STRANDEDNESS:
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PROTEIN
 US-08-466-120-2

Query Match 55.4% Score 1281.5 DB 2 Length 433
 Best Local Similarity 58.7% Pred. No. 3,76-116
 Matches 266 Conservative 52 Mismatches 92 Indels 43 Gaps 7

QY 4 P1SGLTVPVAGNACQNGTSAISLTKFPGGTTPPGAS FSSSSATVIVILEPE 57
 DB 14 POSAVELMKPFGAGWGSQAGSSQTIFFFAVMPHSAGTAVGLAEAFALLTPAEP 73
 QY 59 PPEP-----YPRKQAPVPLGHPVPGVQKVASCHYVNVLSQEGKPFPSVWGACNE 112
 DB 74 SETELPRPKRKKYAVPMQNSVVDWASPHYVNISSCKKEKFPSSVIRKQ-AM 131
 QY 114 YATPNSSTYAMFVSKPKTQVQVPCFCHAVMPFQVLSSEALYPRKQVQVQVQVQV 172

DB 132 YCHESGERTWNTVMPKPCQ-----FVISEQITPKETKQPFQVHATS 177
 QY 174 PDAASSQHPAASPTSPASSQSGSSTSTTGTAAQPMQGVAASTQNNKSSQVQK 333
 DB 178 LPPRASS-PPQITP-----GIPNQLAMIFKIVPAQOQNNKSSQDRLR 220
 QY 234 VLVNIIDHMGVGRAGGQVPAHIFLAIVYFVGAQVNTFTLQCEFTGVALFAS 363
 DB 221 VIVWVADPHSKRQORFAHFTELAVSQEIVDAKQULPGFLOLSKROUALKLK 280
 QY 294 TELMLQVARKYHEDCTLELDEFYKKEFHAPAGVETINIFERSAMPPLCITON 353
 DB 281 IEVMLVETSMKPPSSSEITLTKFSYRPPFAKAGVETINTIFERSAMNLOIND 340
 QY 354 EVALITAINIFSAHPNPVQVPSRPVAFQDPVEALLSTPEKPGDOLPEPMKIVSL 413
 DB 441 TITAINIFSAHPNPVQDQIVPRIGHTYVFAIHAYVSIHHHIDLMFEMKLKLSL 400
 QY 414 PTLSSVSHQVPAFLRQDKKLPPLSLIWDYHE 446
 DB 401 PTLSSVSHQVPAFLRQDKKLPPLSLIWDYHE 433

RESULT 15
 PCT-US94-07266-2
 Sequence 2: Application PCT/US9407266
 GENERAL INFORMATION:
 APPLICANT: CAO, ET AL.
 TITLE OF INVENTION: Peptideic Acid Receptor Epsilon
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CARELLA, RYNE, BAIN, GILFILLAN,
 STREET: 6 BECKER FARM ROAD
 CITY: ROSHLAND
 STATE: NEW JERSEY
 COUNTRY: USA
 ZIP: 07068
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/07266
 FILING DATE: Concurrently
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: FERRARO, GREGORY D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/AGENT NUMBER: 325900-125
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 433 AMINO ACIDS
 TYPE: AMINO ACID
 STRANDEDNESS:
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PROTEIN
 PCT-US94-07266-2

Query Match 55.4% Score 1291.5 DB 5 Length 433
 Best Local Similarity 58.7% Pred. No. 3,76-116
 Matches 266 Conservative 52 Mismatches 92 Indels 43 Gaps 7

QY 4 P1SGLTVPVAGNACQNGTSAISLTKFPGGTTPPGAS FSSSSATVIVILEPE 57

00 14 PDSAVELMKRPAQJAFQSAJGSSST11PEFAPMPSACATACVBI PAAPPTALITPAEPP 73
 0Y 58 DEPPER-----KKKRIATKMI GHELTWGOJKNASHHNVLSLQKGFPPRSVHGAGR 113
 00 74 SHEPTEIRPPKRNKCAAFKMIENLSSVGGNAGSHHNVLSLQKKEKRSVTKG--AA 131
 0Y 114 YACRQSGTQUMDAVMBRKQJQJGJRKCNKAMMBKQVLSSEQJRKRLUKQOQOQPPPS 173
 00 132 YICHSRGHCPNNITYMKRKCQ-----KCVLSFEQJRLKKIKKQHEEQYHATS 177
 0Y 174 EPDASSQJPDASSQJCPASSQJSPSPQJQJTAQPLMUGJVAAGJQJNRPSSQPK 233
 00 178 LPPRASS-PPQJTH-----QJNPEJQJMTFKI VPAQOQJNRPSSQJPLP 220
 0Y 234 VPPMPLADPQSGPMAGAPYAHPTTATISVQJTVPAKQVSPQJQJSPEDQJATKAST 293
 00 221 VPPMPLADPQSGPMAGAPYAHPTTATISVQJTVPAKQJPEPIQJSPEDQJATLKISA 280
 0Y 294 TEIMJLTAPRYNNHTPTPTPKHFTYSKQVPIBATEQVFFNDIPEESPAMPPLGLDPA 353
 00 281 TEVMI VETSRRYNNQSESTTEFKQFSYNNKRPKAKJQJQVFFINTEFPSSKAMNELQJNDP 340
 0Y 354 EYALLIATINPNSADQVNGYPSVFPALQJYVFPALLSYTPKRPQJQJPEPMMKRLVSL 413
 00 341 EFPALLTALISTSAQBPVQJQJQVFPKJQJTYVFPALDQVVSITHPHQJTEPPEMIMKLVSL 400
 0Y 414 KJLSSVHSQJVFALRLQJRKJLPTJLSEIMQVHE 446
 00 401 KJLSSVHSQJVFALRLQJRKJLPTJLSEIMQVHE 433

Search completed: April 25, 2001, 09:56:16
 Job Time: 48 sec

GenCore version 4.5
Copyright (c) 1994 - 2000 Computer Ltd.

4M protein - protein search, using SW model

Run on: April 25, 2001, 00:55:28 ; Search time 19.18 seconds
(without alignment)

1998 041 Millions cell updates/sec

Title: US-09-365-576-3

Percent score: 29.12

Sequence: 1 MSPTSSSTITTPVMSNPSP

Scoring table: MUSUM62

Gapop 10.0, Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 260000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: PIR1*

2: PIR2*

3: PIR3*

4: PIR4*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	2309	99.9	446	2	149021		retinoid X receptor
2	2294	98.8	446	2	150264		retinoid X receptor
3	2033.5	88.0	461	2	104014		steroid hormone re
4	1395.5	60.4	447	2	148975		nuclear orphan rec
5	1379.5	57.7	445	2	156043		steroid hormone re
6	708.5	40.6	878	2	111055		retinoid X receptor
7	704	40.4	936	2	156590		retinoid X receptor
8	638.5	37.4	464	2	146018		retinoid X receptor
9	633.5	37.4	469	2	156919		retinoid X receptor
10	582	32.5	425	2	157316		retinoid X receptor
11	576.5	31.9	429	2	157229		retinoid X receptor
12	573.5	31.8	429	2	151977		retinoid X receptor
13	571	31.7	462	2	152941		retinoid X receptor
14	563	31.4	462	2	156566		retinoid X receptor
15	557	31.1	459	2	156122		retinoid X receptor
16	556.5	31.1	454	2	153433		retinoid X receptor
17	556.5	31.1	458	2	144714		retinoid X receptor
18	555.5	31.0	956	4	156045		retinoid X receptor
19	555	31.0	886	4	151197		retinoid X receptor
20	551.5	30.9	443	2	151991		retinoid X receptor
21	549	30.7	464	2	156558		retinoid X receptor
22	548	30.7	467	2	156558		retinoid X receptor
23	546	30.4	423	2	151761		retinoid X receptor
24	546	30.4	423	2	151761		retinoid X receptor
25	545.5	30.4	438	2	156451		retinoid X receptor
26	545	30.3	452	2	157481		retinoid X receptor
27	541.5	29.5	476	2	151761		retinoid X receptor
28	512	26.4	454	2	156124		retinoid X receptor
29	540	26.4	453	2	156674		retinoid X receptor

ALIGNMENTS

Query Match	99.9%	Score 2309	DB 2:	Length 446
Host local similarity	99.8%	Pred. No. 7,061,581		
Matches	445	Unidirectional	0	Indels 0
1	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
2	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
3	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
4	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
5	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
6	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
7	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
8	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
9	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
10	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
11	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
12	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
13	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
14	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
15	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
16	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
17	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
18	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
19	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
20	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
21	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
22	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
23	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
24	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
25	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
26	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
27	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
28	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446
29	MSPTSSSTITTPVMSNPSP	1	MSPTSSSTITTPVMSNPSP	446

Db 461 INFSARPVNDVDSVEALQDPVEALLSYTRKRVQDLRFPEMLMKLVSLPTLSVH 420
 QY 421 SPVFAIRKQKKIPQISETIMVH 446
 Db 421 SPVFAIRKQKKIPQISETIMVH 446

RESULT 2

159454

orphan nuclear receptor OK 1 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999

C:Accession: 159454; 159464

C:Title: M. Kemat, F. J. J. G. M. Kestrom, A. C. Poltronniko, M. Gustafsson, A.

Proc. Natl. Acad. Sci. U.S.A. 92: 2086-2090, 1995

A:Title: OK-1, a member of the nuclear receptor superfamily that interacts with the 5'-c

A:Reference number: 159454, MIM:95062154

A:Accession: 159454

A:Status: Translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Cross-references: EMBL:U0099; NID:4665941; PIRN:AAA6522.1; PIF:4665942

A:Supplemental source: Splice Registry, hepatic

R:Song, C.; Kokontis, J.M.; Mupakka, R.A.; Rao, S.

Proc. Natl. Acad. Sci. U.S.A. 91: 10809-10813, 1994

A:Title: Ubiquitous receptor: a receptor that modulates gene activation by retinoic acid

A:Reference number: 159464; MIM:95062154

A:Accession: 159464

A:Status: Translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Cross-references: EMBL:U04543; NID:4656617; PIRN:AAA5246.1; PIF:465662

A:Supplemental source: void

C:Keywords: unassigned cDNA related proteins: cDNA transforming protein homology

C:Keywords: DNA binding: zinc finger

F:76-466/2main: cDNA -- inserting protein homology -- cDNA

Query Match

Host Local Similarity 98.8%; Score 2283; DB 2; Length 446;

Matches 400; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSSPSSIPPPVNSGSPSTASPTIKRQVGHIDPPSGSSAVIVILPEDEP 60
 Db 1 MSSPSSIPPPVNSGSPSTASPTIKRQVGHIDPPSGSSAVIVILPEDEP 60
 QY 61 EPRKKRPARKMIGHPVWCTKASPIYNYSPGCGEPFPPSVHGGAPVAPGSG 120
 Db 61 EPRKKRPARKMIGHPVWCTKASPIYNYSPGCGEPFPPSVHGGAPVAPGSG 120
 QY 121 TQGMALPRKQGLVPIPKRFAEMRQVLSPEQIPKRITQKQDQDPPPEPAASS 180
 Db 121 TQGMALPRKQGLVPIPKRFAEMRQVLSPEQIPKRITQKQDQDPPPEPAASS 180
 QY 181 GPRAASTSPASSGSGSGGQGLTAADLMIGLVAAQIQNKSTFSDQKRVWPLG 240
 Db 181 GPRAASTSPASSGSGSGGQGLTAADLMIGLVAAQIQNKSTFSDQKRVWPLG 240
 QY 241 ADQSPAP 300
 Db 241 ADQSPAP 300
 QY 301 TARKYNETETITELKIPYSKDDEHANTLVFINDPFPSPAMPPLCLIDAEVALLIA 360
 Db 301 TARKYNETETITELKIPYSKDDEHANTLVFINDPFPSPAMPPLCLIDAEVALLIA 360
 QY 401 INFSARVNDVDSVEALQDPVEALLSYTRKRVQDLRFPEMLMKLVSLPTLSVH 420
 Db 401 INFSARVNDVDSVEALQDPVEALLSYTRKRVQDLRFPEMLMKLVSLPTLSVH 420
 QY 421 SPVFAIRKQKKIPQISETIMVH 446
 Db 421 SPVFAIRKQKKIPQISETIMVH 446

RESULT 3

JC4014

steroid hormone nuclear receptor NR1 - human

C:Species: Homo sapiens (man)

C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 24-Sep-1999

C:Accession: JC4014

C:Title: NR1, a new member of the gene family encoding the human steroid hormone nu

Rushinar, D.M.; Endo, N.; Rulledge, S.J.; Vogel, K.; Rodan, G.A.; Schmidt, A.

Gene 147, 273-276, 1994

A:Title: NR1, a new member of the gene family encoding the human steroid hormone nu

A:Reference number: JC4014; MIM:9501528

A:Accession: JC4014

A:Molecule type: mRNA

A:Residues: 143; SH1

A:Cross-references: GB:007132; NID:4641961; PIRN:AAA61783.1; PIF:4641962

A:Experimental source: osteosarcoma cells SMS Z/R10

C:Genetics:

A:Gene: GDB:DNK

A:Cross-references: GDB:389570; OMIM:600380

A:Map position: 19q13.3-19q13.4

C:Supplemental source: unassigned cDNA related proteins: cDNA transforming protein homology

F:85-301/2main: cDNA transforming protein homology - EPRK-

F:87-154/2main: DNA binding #status predicted: RIN.

Query Match

Host Local Similarity 88.0%; Score 2043.5; DB 2; Length 401;

Matches 402; Conservative 17; Mismatches 27; Indels 15; Gaps 5;

QY 1 MSSPSSIPPPVNSGSPSTASPTIKRQVGHIDPPSGSSAVIVILPEDEP 51
 Db 1 MSSPSSIPPPVNSGSPSTASPTIKRQVGHIDPPSGSSAVIVILPEDEP 60
 QY 52 VIVPQVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPP 111
 Db 52 VIVPQVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPP 120
 QY 112 GRYVAGSGTQGMADAPPPVQVCPVPRKFAEMRQVLSPEQIPKRITQKQDQ 167
 Db 112 GRYVAGSGTQGMADAPPPVQVCPVPRKFAEMRQVLSPEQIPKRITQKQDQ 180
 QY 181 GPRAASTSPASSGSGSGGQGLTAADLMIGLVAAQIQNKSTFSDQKRVWPLG 240
 Db 181 GPRAASTSPASSGSGSGGQGLTAADLMIGLVAAQIQNKSTFSDQKRVWPLG 240
 QY 241 EPRKKRPARKMIGHPVWCTKASPIYNYSPGCGEPFPPSVHGGAPVAPGSG 300
 Db 241 EPRKKRPARKMIGHPVWCTKASPIYNYSPGCGEPFPPSVHGGAPVAPGSG 300
 QY 301 TARKYNETETITELKIPYSKDDEHANTLVFINDPFPSPAMPPLCLIDAEVALLIA 360
 Db 301 TARKYNETETITELKIPYSKDDEHANTLVFINDPFPSPAMPPLCLIDAEVALLIA 360
 QY 361 INFSARVNDVDSVEALQDPVEALLSYTRKRVQDLRFPEMLMKLVSLPTLSVH 420
 Db 361 INFSARVNDVDSVEALQDPVEALLSYTRKRVQDLRFPEMLMKLVSLPTLSVH 420
 QY 421 SPVFAIRKQKKIPQISETIMVH 446
 Db 421 SPVFAIRKQKKIPQISETIMVH 446

RESULT 4

138975

orphan nuclear receptor LXR-alpha - human

C:Species: Homo sapiens (man)

C:Date: 16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 20-Sep-1999

C:Accession: 138975

C:Title: P. J. M. J. Evans, E. M. J. Evans, E. M. J. Evans, R. A. J. M. J. Evans, D. J.

Genes Dev. 9, 1043-1045, 1995

A:Title: LXR, a nuclear receptor that defines a distinct retinoid response pathway.

476 gvalglaaleuethuiprototyvalthualaleuethuiprototyvalthua 332
 502 tcttctacgctctctacgctctctacgctctctacgctctctacgctctctac 771
 493 tcttctacgctctctacgctctctacgctctctacgctctctacgctctctac 409
 552 ctatct 601
 410 lvalserialatthleuseretvalthiisecolnivalatthleuseretval 426
 602 ct 651
 426 lalserialatthleuseretvalthiisecolnivalatthleuseretval 441
 652 ctatct 701
 442 tttaspvalthiisecolnivalatthleuseretvalthiisecolnival 446
 702 tttaspvalthiisecolnivalatthleuseretvalthiisecolnival 716
 seq_name: ab_09179:HE878950

seq_documentation_block: 957 bp mRNA EST 20-01-2000
 LOCUS HE878950 503492541 REFSEQ_63 Homo sapiens cDNA clone M868.1094376 3
 mRNA sequence:
 HE878950.1 0110327726
 EST
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
 1 (bases 1 to 957)
 REFERENCE
 NIH-MC1 h197269292 nih.gov/
 TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (401) 496-1550
 Email: Robert.Strausberg@nih.gov
 Funding: Department of Energy, contract # DE-AC05-94OR21400
 cDNA library preparation: Life Technologies, Inc.
 DNA sequencing by: Incyte Genomics, Inc.
 clone distribution: MGC clone distribution information can be
 found through the L.M.A.G.E. Consortium/BLM at:
 http://lmage.llnwd.net
 BLAST: LMAN685 row 1 column 14
 High quality sequence stop 746
 Location/Qualifiers

FEATURES
 source
 1..957
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3894996"
 /clone_id="NIH_MGC_69"
 /issue_type="large cell cDNA library, undifferentiated"
 /lab_host="Dh10b (phage-resistant)"
 /note="Organ: lung; Vector: pRW-Sp676; Site: 1; Note:
 Site 2: 5'UTR cloned independently. Primer: oligo 3T.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."
 MAP_COUNT 201 3 335 6 262 4 157 1 1-oligo
 ORIGIN

alignment_scores:
 quality: 119.00 Length: 224
 ratio: 4.946 Gaps: 0
 Percent Similarity: 99.554 Percent Identity: 98.661
 alignment_block:
 US-09-365-576-3 x HE878950

Align seq 1/1 to: HE878950 from 1 to 957

221 tttct 237
 4 ct 52
 227 tttct 254
 53 ct 102
 254 tttct 270
 103 ct 152
 271 tttct 287
 153 ct 202
 287 tttct 304
 203 ct 252
 304 tttct 320
 203 ct 302
 321 tttct 337
 303 ct 352
 337 tttct 354
 353 ct 402
 354 tttct 370
 403 ct 452
 371 tttct 387
 371 ct 402
 403 tttct 404
 403 ct 552
 403 tttct 426
 403 ct 602
 421 tttct 437
 403 ct 651
 437 tttct 444
 437 ct 673
 seq_name: ab_09179:AL169222

seq_documentation_block: 673 bp mRNA EST 31-JAN-1999
 LOCUS AL169222 55725057 RefSeq_63 Homo sapiens cDNA clone
 R18P01 3' end, mRNA sequence:
 AL169222
 VERSION
 K18P01 3' end, mRNA sequence:
 EST
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Muridae; Murinae;
 Rattus.

265 1uc1u11+ValAspPheAlaValsglnValProglyPheLeuGlnLeuGly 281
 551 AAGAGATCGGAGACTTCCTAAACCAAGTCCCTGCTTCTGTAACCTGGG 600
 282 ATGCGTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 298
 601 CCGGAG 650
 298 ATGCGTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 415
 651 TTAG 698
 315 GATGAG 745
 699 TTAG 740

seq_name: qb_est7.1:BE382387

seq_documentation_block:

LOCUS BE382387 721 bp mRNA EST 21-JUL-2000
 DEFINITION: 5312966381 NIH_MGC_19 Homo sapiens: LIMA clone IMAGE:362894 57

ACCESSION BE382387

VERSION BE382387.1 01-04-27752

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 721)

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library: Tissue Culture, Lung, Homo Sapiens, Cerebral Cortex

cDNA Library Arrayed by: The L.M.A.G.E. Consortium (LIMC)

Library: Tissue Culture, Lung, Homo Sapiens, Cerebral Cortex

Clone Distribution: MGC clone distribution information can be

found through the L.M.A.G.E. Consortium/LLMC at: image.llmc.org

Plate: LHM314 Row: 1 Column: 23

High quality sequence: 674

Location/Qualifiers

1..721

Organism "Homo sapiens"

Library "LIMA362894"

Clone "NIH_MGC_19"

Tissue "Lung"

Library "NIH_MGC_19"

Library "NIH_MGC_19"

Library "NIH_MGC_19"

Library "NIH_MGC_19"

Align: 1/1 100 BE382387 1 EST: 1 100 721

254 AlahsPheThnGluLeuAlaValIleSerValIleValIleValAspPhe 270

3 GCGGAG 52

270 GATGAG 287

53 CCGTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 102

287 ATGCGTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404

103 CCGTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 252

304 ATGCTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 320

153 CCGTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 202

420 ATGCTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 437

203 CCGGAG 252

437 ATGCTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 453

254 CCGTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302

354 ATGCTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 370

403 GATGAG 452

470 ATGCTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487

453 CCGGAG 402

487 ATGCTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 403

403 CCGGAG 452

403 CCGGAG 420

453 CCGGAG 502

420 ATGCTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 437

503 CCGGAG 552

437 ATGCTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446

573 TCGTAAAGTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581

seq_name: qb_est7.1:BE391856

seq_documentation_block:

LOCUS BE391856 779 bp mRNA EST 21-JUL-2000
 DEFINITION: 5312966381 NIH_MGC_19 Homo sapiens: LIMA clone IMAGE:361514 57

ACCESSION BE391856

VERSION BE391856.1 01-04-27721

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 779)

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library: Tissue Culture, Lung, Homo Sapiens, Cerebral Cortex

alignment_scores: quality: 964.090 length: 193

Ratio: 4.955 gaps: 0

Percent similarity: 100.000 Percent identity: 98.964

alignment_block:

us-09-365-576-3 8 9942397

us-09-365-576-3

us-09-365-576-3


```
|||||
452 AGTCTCCACAGCTTCTTACGACAGCTTTGGCCACTTCACTGAGCTATGCT 403
261 TTCTTCTGCTGATCTGCTGCTTCTGATCTGCTGCTGCTGCTGCTGCTGCT 277
|||||
402 ATGATGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 353
277 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 294
|||||
352 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 303
294 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 310
|||||
302 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 253
311 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 427
|||||
252 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 203
327 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
|||||
202 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 153
344 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
|||||
152 CTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 103
361 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 477
|||||
102 ATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 53
377 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 494
|||||
52 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4
394 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 494
|||||
2 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2001, 14:15:27 ; Search time 141.49 seconds

(without alignments)
204.151 Million cell updates/sec

Title: US-09-365-576-7

Perfect score: 1860

Sequence: 1 GCGAGGCGACACAGTCGGA.....AAAAAAAAAAAAAAAAAAAA 1860

Scoring table: IDENTIFY_NUC
Gapop 10.0 , Gapov 1.0

Scorptool: 9623517 seqs, 74081774 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008
Lasting first 45 summaries

Database :

EST:*

1: qb_est1: *
2: qb_est2: *
3: qb_est3: *
4: qb_est4: *
5: qb_est5: *
6: qb_est6: *
7: qb_est7: *
8: qb_est8: *
9: qb_est9: *
10: qb_est10: *
11: qb_est11: *
12: qb_est12: *
13: qb_est13: *
14: qb_est14: *
15: qb_est15: *
16: qb_est16: *
17: qb_est17: *
18: qb_est18: *
19: qb_est19: *
20: qb_est20: *
21: qb_est21: *
22: qb_est22: *
23: qb_est23: *
24: qb_est24: *
25: qb_est25: *
26: qb_est26: *
27: qb_est27: *
28: qb_est28: *
29: qb_est29: *
30: qb_est30: *
31: qb_est31: *
32: qb_est32: *
33: qb_est33: *
34: qb_est34: *
35: qb_est35: *
36: qb_est36: *
37: qb_est37: *
38: qb_est38: *
39: qb_est39: *
40: qb_est40: *
41: qb_est41: *
42: qb_est42: *
43: qb_est43: *

44: em_esthum10: *
45: em_esthum11: *
46: em_esthum12: *
47: em_esthum13: *
48: em_esthum14: *
49: em_esthum15: *
50: em_esthum16: *
51: em_esthum17: *
52: em_esthum18: *
53: em_esthum19: *
54: em_esthum20: *
55: em_esthum21: *
56: em_esthum22: *
57: em_esthum23: *
58: em_esthum24: *
59: em_esthum25: *
60: em_esthum26: *
61: em_esthum27: *
62: em_esthum28: *
63: em_est101: *
64: em_est102: *
65: em_est103: *
66: em_est104: *
67: em_est105: *
68: em_est106: *
69: em_est107: *
70: em_est108: *
71: em_est109: *
72: em_est110: *
73: em_est111: *
74: em_est112: *
75: em_est113: *
76: em_est114: *
77: em_est115: *
78: em_est116: *
79: em_est117: *
80: em_est118: *
81: em_est119: *
82: em_est120: *
83: em_est121: *
84: em_est122: *
85: em_est123: *
86: em_est124: *
87: em_est125: *
88: em_est126: *
89: em_est127: *
90: em_est128: *
91: em_est129: *
92: em_est130: *
93: em_est131: *
94: em_est132: *
95: em_est133: *
96: em_est134: *
97: em_est135: *
98: em_est136: *
99: em_est137: *
100: em_est138: *
101: em_est139: *
102: em_est140: *
103: em_est141: *
104: em_est142: *
105: em_est143: *
106: em_est144: *
107: em_est145: *
108: em_est146: *
109: em_est147: *
110: em_est148: *
111: em_est149: *
112: em_est150: *
113: em_est151: *
114: em_est152: *
115: em_est153: *
116: em_est154: *

117: qb_est48:*
 118: qb_est49:*
 119: qb_est50:*
 120: qb_est51:*
 121: qb_est52:*
 122: qb_est53:*
 123: qb_est54:*
 124: qb_est55:*
 125: qb_est56:*
 126: qb_est57:*
 127: qb_est58:*
 128: qb_est59:*
 129: qb_est60:*
 130: qb_est61:*
 131: qb_est62:*
 132: qb_est63:*
 133: qb_est64:*
 134: qb_est65:*
 135: qb_est66:*
 136: qb_est67:*
 137: qb_est68:*
 138: qb_est69:*
 139: qb_est70:*
 140: qb_est71:*
 141: qb_est72:*
 142: qb_est73:*
 143: qb_est74:*
 144: qb_est75:*
 145: qb_est76:*
 146: qb_est77:*
 147: qb_est78:*
 148: qb_est79:*
 149: qb_est80:*
 150: qb_est81:*
 151: qb_est82:*
 152: qb_est83:*
 153: qb_est84:*
 154: qb_est85:*
 155: qb_est86:*
 156: qb_est87:*
 157: qb_est88:*
 158: qb_est89:*
 159: qb_est90:*
 160: qb_est91:*
 161: qb_est92:*
 162: qb_est93:*
 163: qb_est94:*
 164: qb_est95:*
 165: qb_est96:*
 166: qb_est97:*
 167: qb_est98:*
 168: qb_est99:*
 169: qb_est100:*
 170: qb_est101:*
 171: qb_est102:*
 172: qb_est103:*
 173: qb_est104:*
 174: qb_est105:*
 175: qb_est106:*
 176: qb_est107:*
 177: qb_est108:*
 178: qb_est109:*
 179: qb_est110:*
 180: qb_est111:*
 181: qb_est112:*
 182: qb_est113:*
 183: qb_est114:*
 184: qb_est115:*
 185: qb_est116:*
 186: qb_est117:*
 187: qb_est118:*
 188: qb_est119:*
 189: qb_est120:*

190: em_qss_pln1:*
 191: em_qss_pln2:*
 192: em_qss_pro:*
 193: em_qss_rpd1:*
 194: em_qss_rpd2:*
 195: em_qss_rpd3:*
 196: em_qss_rpd4:*
 197: em_qss_rpd5:*
 198: em_qss_vrt1:*
 199: em_qss_vrt2:*
 200: em_qss_vrt3:*
 201: qb_qss1:*
 202: qb_qss2:*
 203: qb_qss3:*
 204: qb_qss4:*
 205: qb_qss5:*
 206: qb_qss6:*
 207: qb_qss7:*
 208: qb_qss8:*
 209: qb_qss9:*
 210: qb_qss10:*
 211: qb_qss11:*
 212: qb_qss12:*
 213: qb_qss13:*
 214: qb_qss14:*
 215: qb_qss15:*
 216: qb_qss16:*
 217: qb_qss17:*
 218: qb_qss18:*
 219: qb_qss19:*
 220: qb_qss20:*
 221: qb_qss21:*
 222: qb_qss22:*
 223: qb_qss23:*
 224: qb_qss24:*
 225: qb_qss25:*
 226: qb_qss26:*
 227: qb_qss27:*
 228: qb_qss28:*
 229: qb_qss29:*
 230: qb_qss30:*
 231: qb_qss31:*
 232: qb_qss32:*
 233: qb_qss33:*
 234: qb_qss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB	ID	Description
1	736	39.6	757	172	BC067766 BC05711
2	681.8	36.7	788	141	BE915476 BE915476
3	658.6	35.4	725	21	AF528699 AF528699
4	651.6	35.0	667	112	AW231122 AW231122
5	650.4	35.0	706	18	AF1322708 AF1322708
6	627.2	33.7	743	143	BF099698 BF099698
7	627.2	33.3	627	121	AW911822 AW911822
8	618.4	33.2	620	148	BF468445 BF468445
9	618.4	33.0	660	144	BF116819 BF116819
10	613	33.0	759	17	AF194859 AF194859
11	604.8	32.5	704	140	BF847826 BF847826
12	596	32.0	667	13	AF920102 AF920102
13	591	31.8	675	17	AF169222 AF169222
14	587.2	31.6	687	112	AW208542 AW208542
15	583.4	31.4	807	149	BF536092 BF536092
16	580	31.2	624	10	AA646678 AA646678
17	569.8	30.6	957	140	BF878950 BF878950
18	566.6	30.5	641	4	AA222594 AA222594

19	562	30	579	167	479414	BE33534	50122422
20	568	30	619	170	A140494	A140494	E3122780
21	566	28	619	170	PF06019	PF06019	U01007
22	566	28	619	170	PF06019	PF06019	U01007
23	520	28	619	170	PF06019	PF06019	U01007
24	512	27	619	170	PF06019	PF06019	U01007
25	511	27	619	170	PF06019	PF06019	U01007
26	511	27	619	170	PF06019	PF06019	U01007
27	509	27	619	170	PF06019	PF06019	U01007
28	509	27	619	170	PF06019	PF06019	U01007
29	509	27	619	170	PF06019	PF06019	U01007
30	499	26	619	170	PF06019	PF06019	U01007
31	495	26	619	170	PF06019	PF06019	U01007
32	494	26	619	170	PF06019	PF06019	U01007
33	492	26	619	170	PF06019	PF06019	U01007
34	491	26	619	170	PF06019	PF06019	U01007
35	489	26	619	170	PF06019	PF06019	U01007
36	486	26	619	170	PF06019	PF06019	U01007
37	481	25	619	170	PF06019	PF06019	U01007
38	481	25	619	170	PF06019	PF06019	U01007
39	479	25	619	170	PF06019	PF06019	U01007
40	478	25	619	170	PF06019	PF06019	U01007
41	478	25	619	170	PF06019	PF06019	U01007
42	473	25	619	170	PF06019	PF06019	U01007
43	470	25	619	170	PF06019	PF06019	U01007
44	470	25	619	170	PF06019	PF06019	U01007
45	469	25	619	170	PF06019	PF06019	U01007

ALIGNMENTS

RESULT 1
LOCUS BG067766/c
DEFINITION H3057H11-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
ACCESSION BG067766
VERSION 1
KEYWORDS EST
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE 1
AUTHORS Kargul, G. J., Dudekula, D. B., Qian, Y., Lim, M. K., Jaradat, S. A., Tanaka, T. S., Gail, M. G., and R. M. S. H.
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other ESTs: H3057H11-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: ckargul@nia.nih.gov
This clone set has been freely distributed to the community. Please visit <http://www.nia.nih.gov/ncic/tm/15k> for details.
Seq primer: 21m3 Forward
Place: H3057 Row: H Column: 11
High quality sequence stop: 757
PolyA=yes

FEATURES

Source
1..757
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/note="H3057H11"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_str="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="Vector: pSP6B1; Size: 1.1 kb; Salt: 0.5 M; NoT: 10"

19	562	30	579	167	479414	BE33534	50122422
20	568	30	619	170	A140494	A140494	E3122780
21	566	28	619	170	PF06019	PF06019	U01007
22	566	28	619	170	PF06019	PF06019	U01007
23	520	28	619	170	PF06019	PF06019	U01007
24	512	27	619	170	PF06019	PF06019	U01007
25	511	27	619	170	PF06019	PF06019	U01007
26	511	27	619	170	PF06019	PF06019	U01007
27	509	27	619	170	PF06019	PF06019	U01007
28	509	27	619	170	PF06019	PF06019	U01007
29	509	27	619	170	PF06019	PF06019	U01007
30	499	26	619	170	PF06019	PF06019	U01007
31	495	26	619	170	PF06019	PF06019	U01007
32	494	26	619	170	PF06019	PF06019	U01007
33	492	26	619	170	PF06019	PF06019	U01007
34	491	26	619	170	PF06019	PF06019	U01007
35	489	26	619	170	PF06019	PF06019	U01007
36	486	26	619	170	PF06019	PF06019	U01007
37	481	25	619	170	PF06019	PF06019	U01007
38	481	25	619	170	PF06019	PF06019	U01007
39	479	25	619	170	PF06019	PF06019	U01007
40	478	25	619	170	PF06019	PF06019	U01007
41	478	25	619	170	PF06019	PF06019	U01007
42	473	25	619	170	PF06019	PF06019	U01007
43	470	25	619	170	PF06019	PF06019	U01007
44	470	25	619	170	PF06019	PF06019	U01007
45	469	25	619	170	PF06019	PF06019	U01007

Db 657 CTCCTCTGCTTTTATTAATAAAAAACAATAAAAA 652

RESULT 6
BF099698 743 bp mRNA EST 19-061-2900
LOCUS BF099698
DEFINITION Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus musculus cDNA clone IMAGE:3979183 5'.

ACCESSION BF099698
VERSION BF099698.1 GI:10882224
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Fukuyama: Metazoa: Chordata: Granulata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus (bases 1 to 743)
NIH-MGC built 7/99/01 rib-99/01
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The J.M.A.C.L. Consortium (J.M.C.)
DNA Sequencing by: Invitrogen Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the J.M.A.C.L. Consortium/BLAST at:
<http://image.llnl.gov>
plate: HLM9172 row: m column: 08
High quality sequence stop: 700
location/qualifiers
1..743

FEATURES
Source
organism: "Mus musculus"
strain: "FVB/N"
cell_type: "fibroblast"
vector: "pBluescript II"
clone_lib: "NCI-CGAP_Mam"
tissue_type: "tumor, biopsy sample"
map: "190101" chr: "X" pos: "111111"
/lab: "FBI/NIH"
/note: "Organ: mammary; Vector: pTZ19, pTZ19, pTZ19; Site: 2; Not: Cloned unidirectionally. Primer: cDNA dt. library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 156 a 235 c 213 g 138 t 1 others

ORIGIN

Query Match 33.7% Score 627.2, 28 113, Length 743.
Best Local Similarity 96.3% Pos: 4, 3e-177
Matches 707; Conservation 0; Mismatches 18; Indels 9; Gaps 9

UY 683 CCTGCTCTGCTTTTATTAATAAAAAACAATAAAAA 741
Db 7 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 66
UY 742 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 801
Db 67 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 126
UY 802 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 861
Db 127 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 186
UY 842 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 941
Db 187 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 246
UY 922 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 981
Db 247 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 306

UY 982 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 1041
Db 407 GGTGCTGCTGCTTTTATTAATAAAAAACAATAAAAA 466
UY 1042 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 1101
Db 367 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 426
UY 1102 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 1261
Db 427 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 486
UY 1162 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 1221
Db 487 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 546
UY 1222 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 1281
Db 547 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 606
UY 1282 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 1342
Db 606 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 666
UY 1342 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 1402
Db 666 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 728
UY 1402 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 1462
Db 728 GTAGATGCTGCTTTTATTAATAAAAAACAATAAAAA 788

RESULT 7
AW911822 527 bp mRNA EST 25 MAY-2 1999
DEFINITION
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus musculus cDNA clone IMAGE:151750 5' similar to GI:509419 Mus musculus retina X receptor interacting protein (MOUSE); mRNA sequence.

ACCESSION AW911822
VERSION AW911822.1 GI:8577405
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Fukuyama: Metazoa: Chordata: Granulata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus (bases 1 to 527)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The J.M.A.C.L. Consortium (J.M.C.)
DNA Sequencing by: Invitrogen Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the J.M.A.C.L. Consortium/BLAST at:
<http://image.llnl.gov>
plate: HLM9172 row: m column: 08
High quality sequence stop: 477.
location/qualifiers
1..527

FEATURES
Source
organism: "Mus musculus"
strain: "FVB/N"
cell_type: "retina"
vector: "pBluescript II"
clone_lib: "NCI-CGAP_Mam"
tissue_type: "tumor, biopsy sample"
map: "190101" chr: "X" pos: "111111"
/lab: "FBI/NIH"
/note: "Organ: mammary; Vector: pTZ19, pTZ19, pTZ19; Site: 2; Not: Cloned unidirectionally. Primer: cDNA dt. library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

DB 690 AAACCTTTTCAAGAAATGAAACCAAGGAGCTAACACCTCCATTTGAGG 738

Search completed: April 29, 2001, 15:34 25
Job time: 4738 sec

Query Match	80.08	Score	711	DB	5	Length	881
Best Local Similarity	46.08	Prod. No.	4,900,52				
Matches	167	Conservative	82	Mismatches	151	Indels	64
				Gaps	13		

[illegible]

00	Tophiaridaceae; Lepitridaceae; Ceratitids.
01	NCBI:38167213;
02	11)
03	SEQUENCE FROM N.A.
04	Veritas M., Mavroudis M., Koulakakis N., Zacharopoulou A., Minizas A.;
05	"Cloning and characterization of the cys250 gene receptor of the
06	motiliferous flat-trailly Ceratitids capitata."
07	Submitted (Feb. 1998) to the PubMed database.
08	CC
09	1. SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
10	EMBL: AJ224341; CCAAT1967.1; -;
11	DR
12	HSSP: P20393; 1A6Y.
13	DR
14	INTERPRO: IPR000324; -;
15	DR
16	INTERPRO: IPR000546; -;
17	DR
18	INTERPRO: IPR001628; -;
19	DR
20	INTERPRO: IPR001723; -;
21	DR
22	INTERPRO: IPR004069; -;
23	DR
24	PFAM: PF00104; hormone_rec; 1.
25	DR
26	PFAM: PF00105; z1-ca; 1.
27	DR
28	PRINTS: PR00047; STROLODINGER.
29	PRINTS: PR00350; VITAMINDR.
30	DR
31	PRINTS: PR00398; STROBROOME.
32	DR
33	PRINTS: PR01283; ECDYSTEROIDR.
34	DR
35	PROSITE: PS00041; NUCLEAR_RECEPTOR; 1.
36	KW
37	Receptor; Transcription regulation; DNA-binding; Nuclear protein;
38	Zinc-finger.
39	FT
40	NON_TER 1 1
41	SEQUENCE 673 AA; 74740 MW; AFB6A517B3BFF2BF CYP64;

[illegible]

00 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo
 01 NCBI_TaxID:9606;
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100

Query Match: 27.79, Score: 641, Length: 484
 Best Local Similarity: 37.0%, Prod. No. 1,66-46;
 Matches: 144; Conservative: 69; Missed: 115; Indels: 92; Gaps: 2;
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100

Query Match: 27.69, Score: 605.5, Length: 484
 Best Local Similarity: 34.3%, Prod. No. 1,76-46;
 Matches: 151; Conservative: 77; Missed: 159; Indels: 53; Gaps: 16;
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100

Search completed: April 25, 2001, 09:59:00
 Job time: 159 sec
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100

PL	Proc. Natl. Acad. Sci. U.S.A.	92:2006-2100(1995).
RM	[2]	
RK	SPOUNCE FROM N.A.	
KC	STRAIN STRAIN-DAMLEY: ISSUE Vertical:	
KA	MELINE-05062154: PubMed:7971966;	
FA	Said G., Kozakian J.M., Elloukha R.A., Liao S.-J.	
RI	"Multigene receptor: a receptor that modulates gene activation by	
RT	retinoic acid and thyroid hormone receptors."	
RL	proc. Natl. Acad. Sci. U.S.A. 91:10809-10814(1994).	
CC	-1 FUNCTION: CYPAN RECEPTOR BINDS INTERMITTENTLY TO POU-SF STRAINED	
CC	OILIONOLECTIDE DIRECT REPAIRS HAVING THE CONSENSUS HALF SITE	
CC	SEQUENCE 5'-AGGTGA-3' AND 4-NI SPACING (DR 4).	
CC	-1 SUBUNIT: FORMS A HETEROOLMER WITH RXR.	
CC	-1 SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).	
CC	-1 SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.	
CC	NH1 SUBFAMILY.	
CC		
CC	This SWISS-PROT entry is copyright . It is produced through a collaboration	
CC	between The Swiss Institute of Bioinformatics and the EMBL institution	
CC	The European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	misleading and this statement is not removed, updated, used, by and for commercial	
CC	entities requires a license agreement from EMBL and/or the EBI.	
CC	or send an email to license@ebi.ac.uk.	
CC		
DZ	EMBL: 029289; AAA9522.1 ;	
DZ	EMBL: 014533; AAA52461.1 ;	
DZ	HSDP: P04372; IHCP.	
DZ	InterPro: IPR00536; ?	
DZ	InterPro: PR001628; ?	
DZ	Pfam: PF00104; hormone_rec_1.	
DZ	Pfam: PF00105; zt-c4_1.	
DZ	PRINTS: PR00047; STROLOPE_NHR.	
DZ	PROSITE: PS00031; NUCLEAR_RECEPTOR_1.	
KW	Receptor; Transcription regulation; DNA-binding; Nuclear protein;	
KW	Zinc-finger.	
FT	DNA_BIND 78 145 NUCLEAR RECEPTOR TYPE.	
FT	ZN_FING 78 98 c4-TYPE.	
FT	ZN_FING 116 140 c4-TYPE.	
FT	E_DOMAIN 216 416 Hormone binding site (HBM1A).	
FT	E_DOMAIN 162 168 POLY-QIN.	
FT	E_DOMAIN 169 172 POLY-PRO.	
FT	CORECTCT 33 43 G -> V (IN REF. 2).	
FT	CORECTCT 52 54 MISSING (IN REF. 2).	
FT	CONFLICT 219 219 A -> V (IN REF. 2).	
SO	SOURCE 446 AA; 49735 MW; 1406462 PROTEIN DATA BANK.	
Query Match	98.8% Score 2284; DB 1; Length 446;	
Best Local Similarity	98.2% Prod. No. 5,40-154;	
Matches 446; Conservation	6; Mismatches 2; Indels 0; Gaps 0;	
QY	I MSSTSLDTLVVGNSSPPSPISALSTPIKEGACDITFGSGSSASATVLLLEDEDE 60	
DB	1 MSSPTSLDPLEPGNSPPSPSTSSPSPIKEGGDIDPFSSSSAVTVLLEDEDE 60	
QY	EKKKKCKPAEKMKCHEICGVCGAKSGEHYNYLSGYTKGFRRSVNHGACRYAPASST 120	
IE	EEPKRKCHATEMLGHEDTVEGCKAKSGHYNYLSGYTKGFRRSVNHGACRYAPASST 120	
QY	TTCOMAIAMPKPKCVLPKPKCFEAMEPGVSTPHPREPTFGVGLGPFPSSVAASS 180	
DB	1 TCOMAIAMPKPKCVLPKPKCFEAMEPGVSTPHPREPTFGVGLGPFPSSVAASS 180	
QY	GTAASPDTSFASASGSSESHLDIAAEFLMDSGVAAGCNKPSFSNGEVTHWHS 240	
DB	181 ARPASGITSFASASGSSESHLDIAAEFLMDSGVAAGCNKPSFSNGEVTHWHS 240	
QY	ATFESHPAGQAFAREELIALSVGCVTFDFGVVETLGQPEFYALNKASTHEMLIG 300	
DB	241 ADPSHDARQGQRAFEHLIALSVGEIVFKQVGVEFDGREGCLALKASTHEMLILE 300	
QY	TAPPYNIEETFTFIKNIFYSKVFHPANTLVEFINIFEFSSMRKLGLDMDEVALLA 360	

[illegible]

```

01 01.08T.996 (Ref. 34, last sequence update)
02 01-JUL-1999 (Ref. 34, last annotation update)
03 ETYVINE RECEPTOR (ETYVINE RECEPTOR) (200-AMINO ACID)
04 RECEPTOR (200 RECEPTOR) (200).
05 ECR OR NR1H1.
06 Chromomus tentans (Midie).
07 Pharyngodon latipes (Midie).
08 Pharyngodon latipes (Midie).
09 Pharyngodon latipes (Midie).
10 Pharyngodon latipes (Midie).
11 Pharyngodon latipes (Midie).
12 Pharyngodon latipes (Midie).
13 Pharyngodon latipes (Midie).
14 Pharyngodon latipes (Midie).
15 Pharyngodon latipes (Midie).
16 Pharyngodon latipes (Midie).
17 Pharyngodon latipes (Midie).
18 Pharyngodon latipes (Midie).
19 Pharyngodon latipes (Midie).
20 Pharyngodon latipes (Midie).
21 Pharyngodon latipes (Midie).
22 Pharyngodon latipes (Midie).
23 Pharyngodon latipes (Midie).
24 Pharyngodon latipes (Midie).
25 Pharyngodon latipes (Midie).
26 Pharyngodon latipes (Midie).
27 Pharyngodon latipes (Midie).
28 Pharyngodon latipes (Midie).
29 Pharyngodon latipes (Midie).
30 Pharyngodon latipes (Midie).
31 Pharyngodon latipes (Midie).
32 Pharyngodon latipes (Midie).
33 Pharyngodon latipes (Midie).
34 Pharyngodon latipes (Midie).
35 Pharyngodon latipes (Midie).
36 Pharyngodon latipes (Midie).
37 Pharyngodon latipes (Midie).
38 Pharyngodon latipes (Midie).
39 Pharyngodon latipes (Midie).
40 Pharyngodon latipes (Midie).
41 Pharyngodon latipes (Midie).
42 Pharyngodon latipes (Midie).
43 Pharyngodon latipes (Midie).
44 Pharyngodon latipes (Midie).
45 Pharyngodon latipes (Midie).
46 Pharyngodon latipes (Midie).
47 Pharyngodon latipes (Midie).
48 Pharyngodon latipes (Midie).
49 Pharyngodon latipes (Midie).
50 Pharyngodon latipes (Midie).
51 Pharyngodon latipes (Midie).
52 Pharyngodon latipes (Midie).
53 Pharyngodon latipes (Midie).
54 Pharyngodon latipes (Midie).
55 Pharyngodon latipes (Midie).
56 Pharyngodon latipes (Midie).
57 Pharyngodon latipes (Midie).
58 Pharyngodon latipes (Midie).
59 Pharyngodon latipes (Midie).
60 Pharyngodon latipes (Midie).
61 Pharyngodon latipes (Midie).
62 Pharyngodon latipes (Midie).
63 Pharyngodon latipes (Midie).
64 Pharyngodon latipes (Midie).
65 Pharyngodon latipes (Midie).
66 Pharyngodon latipes (Midie).
67 Pharyngodon latipes (Midie).
68 Pharyngodon latipes (Midie).
69 Pharyngodon latipes (Midie).
70 Pharyngodon latipes (Midie).
71 Pharyngodon latipes (Midie).
72 Pharyngodon latipes (Midie).
73 Pharyngodon latipes (Midie).
74 Pharyngodon latipes (Midie).
75 Pharyngodon latipes (Midie).
76 Pharyngodon latipes (Midie).
77 Pharyngodon latipes (Midie).
78 Pharyngodon latipes (Midie).
79 Pharyngodon latipes (Midie).
80 Pharyngodon latipes (Midie).
81 Pharyngodon latipes (Midie).
82 Pharyngodon latipes (Midie).
83 Pharyngodon latipes (Midie).
84 Pharyngodon latipes (Midie).
85 Pharyngodon latipes (Midie).
86 Pharyngodon latipes (Midie).
87 Pharyngodon latipes (Midie).
88 Pharyngodon latipes (Midie).
89 Pharyngodon latipes (Midie).
90 Pharyngodon latipes (Midie).
91 Pharyngodon latipes (Midie).
92 Pharyngodon latipes (Midie).
93 Pharyngodon latipes (Midie).
94 Pharyngodon latipes (Midie).
95 Pharyngodon latipes (Midie).
96 Pharyngodon latipes (Midie).
97 Pharyngodon latipes (Midie).
98 Pharyngodon latipes (Midie).
99 Pharyngodon latipes (Midie).
100 Pharyngodon latipes (Midie).

```


•

•

•

08-09-365-576-3 X MM009419

Align seq 1/1 for MM009419 from: 1 to: 1841

1 MetSerSerProThrSerSerLeuAspThrProValProGlyAsnGlySer 17
 143 ATGCTCTGGCCCAATAACTCTCTGTAATAGCTGGCTGGCTAAAGAAATGGCTTC 180
 17 CTTCGCTGTTGCTCTTCTGATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 34
 181 TCT 230
 84 LngLThrAspProProThrLeuGlySerGlyGlySerSerSerAlaIle 50
 241 AGGAAATGATGATCT 280
 51 ValValIleLeuGlyProGlyAspGlyProGlyAsnGlyAsnGly 67
 281 GCTGATCT 330
 67 yProGlyProGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 84
 341 GCTGATCT 380
 84 yValAsnGlyGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 109
 481 AGCT 430
 101 GAGGATCT 117
 441 TCT 480
 117 GCT 134
 481 GCT 530
 144 GCT 150
 541 TCT 580
 151 GCT 167
 581 GCT 630
 167 GCT 184
 641 GCT 680
 184 LngLThrProGlyGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 209
 681 GCT 740
 201 GCT 217
 741 GCT 780
 217 LngLThrProGlyGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 234
 781 GCT 830
 244 GCT 250
 841 GCT 890
 251 GCT 267
 881 GCT 930
 267 GCT 284
 941 GCT 980
 284 GCT 300

941 ACCAGATCGCT 1030
 301 ThrAlaGATCT 317
 1041 ACAGCAGACAGCT 1080
 317 pPheThrIleSerGlyAspAspPheHisArgAlaGlyGlyGlyGlyGlyGly 334
 1081 CTTCACCT 1140
 334 HeLeuAspProLePheGlyIlePheSerArgAlaMetArgArgGlyGlyGly 350
 1131 TCATCATTCATCT 1180
 351 AspAspAlaGlyGlyArgAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 367
 1181 GACCATCT 1240
 367 pArgProAsnValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 484
 1241 TCGCT 1280
 384 yValGlyAlaLeuLeuSerGlyThrArgGlyGlyGlyGlyGlyGlyGlyGly 400
 1281 AGCT 1340
 401 LeuArgPheProArgMetLeuMetGlySerValSerLeuArgGlyGlyGlyGly 417
 1341 CT 1400
 417 pSerValHisSerGlyGlyGlyValPheAlaLeuArgGlyGlyGlyGlyGly 444
 1381 CT 1440
 444 euProLeuLeuSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 466
 1441 TCGCT 1466

seq_name: gb_f02:RNU20489

seq_documentation_block: 1877 bp mRNA 04 Jul-1995
 L/GENS RNU20389 Rat20389 rat20389 rat20389 rat20389 rat20389
 DEFINITION Rat20389 rat20389 rat20389 rat20389 rat20389
 ACCESSION 020389
 VERSION 020389.1 GI:665941
 KEYWORDS
 SOURCE rat
 ORGANISM Rat20389
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurinae; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 1877)
 AUTHORS Teboul,M., Emarek,E., Li,G., Wikstrom,A.C., Petto-Bulko,M. and
 Gustafsson,J.A.
 OR-1, a member of the nuclear receptor superfamily that interacts
 with the 9-cis-retinoic acid receptor
 Proc. Natl. Acad. Sci. U.S.A. 92 (6), 2096-2100 (1995)
 MEDLINE 95199298
 REFERENCE 2 (bases 1 to 1877)
 AUTHORS Emarek,E.
 Direct Submission
 Submitted (NOV-1995) Eva Emarek, Center for Microbiology at
 Novum, Center for Microbiology, Karolinska Institute, Novum,
 Huddinge, S-141 86, Sweden
 Location/Qualifiers

FEATURES

source
 1..1877
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_lib="Jambaa ZAP II cDNA library, Stratagene catalog
 number 946-514"
 /sex="male"

CDS
 /issue_type "river"
 /product_start 1
 /product_end 1
 /protein_id "AAA69522.1"
 /db_xref="db:AA69522.1"

misc_feature
 422..625
 /note="noncoding RNA binding domain"

BASE COUNT 384 a 588 c 544 g 361 t
 ORIGIN

alignment_statistics
 quality: 428.09 Length: 446
 Ratio: 5.121 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.206

alignment_block
 US-09-365-576-3 x NM020489 ..

Align seq 1/1 to: NM020489 from: 1 to: 1877

```

1 MetSerSerProThrSerSerLeuAspThrProValProGlyValSerGly 17
  |||||
191 ATGCTTCTCCCAACAGTCTCTGCAACCTCCCTGCTGCTCCCAATGCTT 240
  |||||
17 PheAlaProSerPheSerAlaThrSerProThrIleGlyGlnGlyGly 34
  |||||
241 TCCGACGCGCAGIAGCTGCTCCACCTTCACCCACATTAATGACGACGAC 290
  |||||
34 IuGluThrAspProThrProGlySerGlyGlySerSerAlaTyrIle 50
  |||||
291 AGGAAATGATGACCTGACAGCTGCTGAAAGGCTGACCTGCTGCTCAAT 340
  |||||
51 ValValIleLeuGluProGluAspGluProGluAlaGlyValGlySerG 67
  |||||
341 GTATCATCTTAAAGCTAGAGGATGAACTGAGGCTGAAAGGAAAGAGG 390
  |||||
67 GTTAAATATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 440
  |||||
391 TCCAGGCGCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 440
  |||||
84 ValAlaSerGlyPheHisTyrAsnValLeuSerGlySerGlyGlyGly 100
  |||||
441 AGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490
  |||||
101 PhePheArgAlaGluSerValIleHisGlyValGlyValGlyValGly 117
  |||||
491 TTCTTGGGCTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
  |||||
117 GglySerGlyTyrGlySerGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 124
  |||||
541 GGGCAGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 590
  |||||
134 GGGCAGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 590
  |||||
591 TGTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640
  |||||
151 LeuSerGluGluGluIleArgGlyValGlyValGlyValGlyValGly 167
  |||||
641 GTTCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 690
  |||||
167 GgluProPheThrSerSerGluAlaSerSerSerGlyAlaProGlu 184
  |||||
691 GCAATCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 740
  |||||

```

```

184 ValAlaSerProGlyThrSerGluAlaSerSerGlySerGlyGlyGly 206
  |||||
741 GAGCTTCTCCCAACAGTCTCTGCAACCTCCCTGCTGCTCCCAATGCTT 790
  |||||
201 GTTAAATATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 217
  |||||
791 TCCAGGCGCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
  |||||
217 ValAlaSerGlyPheHisTyrAsnValLeuSerGlySerGlyGlyGly 234
  |||||
841 TGTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 890
  |||||
234 GTTAAATATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 250
  |||||
891 TGTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 940
  |||||
251 GluThrPheAlaHisPheThrGluGluAlaIleIleSerValGlnGlu 267
  |||||
941 GAAGCTTCTCCCAACAGTCTCTGCAACCTCCCTGCTGCTCCCAATGCTT 990
  |||||
267 ValAlaSerPheAlaHisPheThrGluGluAlaIleIleSerValGln 294
  |||||
991 GTTAAATATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1040
  |||||
294 SpheAlaSerPheAlaHisPheThrGluGluAlaIleIleSerValGln 300
  |||||
1041 ACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1090
  |||||
301 ThrAlaArgAlaGlyTyrAsnHisGluThrGluGlySerIleThrPhe 317
  |||||
1091 ACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
  |||||
317 PheThrTyrSerGlyAspAspPheHisPheAlaGlyValGlnValGlu 334
  |||||
1141 GTTAAATATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1190
  |||||
334 IleIleAsnProIleThrGluProPheSerAlaGluMetArgAlaGlu 350
  |||||
1191 TCAATCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1240
  |||||
351 AspAspAlaGluThrAlaLeuGluAlaIleIleSerPheSerAlaAs 367
  |||||
1241 GAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1290
  |||||
367 PheProAsnValGlnIleProSerArgValGluAlaGlnGlnIlePro 384
  |||||
1291 GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1340
  |||||
384 ValValGluAlaGluLeuSerGlyThrAlaGlySerArgThrGluAsp 400
  |||||
1341 ATGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1390
  |||||
401 LeuArgPhePheArgAlaGluLeuMetIleGlyValLeuSerLeuArg 417
  |||||
1391 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
  |||||
417 PheValHisSerGluGluIleValPheAlaGluAlaGluGlnAspGly 434
  |||||
1441 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1490
  |||||
434 GlnProProLeuLeuSerGluIleIleThrAspValHisSerGlu 446
  |||||
1491 TGTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1540
  |||||

```

seq_name: gb_pat1:AA9139

seq_documentation_block:

Accession: A49139

Definition: Sequence 1 from Patent WO/95/00111

Accession: A49139

Version: A49139.1 GI:2302694

Keywords:

PAT

G7-MAP-1997

SOURCE: Unknown.
 ORGANISM: Unknown.
 REFERENCE: 1 (bases 1 to 1959)
 AUTHORS: Liou, S. and Song, C.
 TITLE: Isolated nucleic acid encoding a ubiquitous nuclear receptor
 JOURNAL: Proc. Natl. Acad. Sci. USA 94:16-A x 17-JUN 1997
 FEATURES: location/qualifiers
 SOURCE: 1..1959
 BASE COUNT: 406 a 608 c 564 g 481 t
 ORIGIN:

alignment_scores:
 quality: 2447.50 length: 446
 Ratio: 5.085 gaps: 1
 Percent Similarity: 99.103 Percent Identity: 97.085

alignment_block:
 US-09-365-576-3 x 146766 ..

Align seq 1/1 to: 146766 from: 1 to: 1959

```

1 MotSerSerProthSerSotLeuAspLhrProValProGlyAsnLysc 17
|||||
256 AIGCTTCCTCCCAACAGTTCCTGCAACCTCTCTCTGCAACAGTTC 405
|||||
17 PProGlyProSerTheSerAlaThrSerProThrTleLysLcLcLc 44
|||||
406 TCGGCACACCACTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 455
|||||
34 InsGlnAspProProProGlySerGlnLysSerSotAlaTyrLc 50
|||||
356 AGGAGCTGATCACTCCAGAGCTCTGAAAGGCTGCAAGCTGCTAC 405
|||||
51 ValValLleLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLc 67
|||||
406 GTG.....CAATCAAGAGATCAACTGAGAGCAAGAGCAAGAG 446
|||||
67 YProAlaProLysMetLcLcLcLcLcLcLcLcLcLcLcLcLcLc 84
|||||
447 TCGGGCTGCAAGAGCTGAGAGCAAGCTGCTGCTGCTGCTGCTG 496
|||||
84 YsaLAspGlyPheMetLysTyrAspValLcLcLcLcLcLcLcLc 100
|||||
497 AGGCTCTGAGGCTTTCCTCACTCAATGCTCTCACTGCTGCAAG 546
|||||
101 PhePheAspArgSerValValHisGlyValAlaGlyValTyrAla 117
|||||
547 TTCTTCTGAGGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596
|||||
117 GcLysSerGlyThrLysGlnMetAspAlaPheMetAspAlaLys 134
|||||
597 GGGCAAGGCAAGCTGGCAGATGCTCTGATGGGCTGCAATGCTG 646
|||||
134 GcGlySerAlaGluArgLysGlySerLcLcLcLcLcLcLcLcLc 150
|||||
647 TCTGCAAGCTGAGTCACTGCAAGAGCTGCTGCAAGAGCAAGT 696
|||||
151 LeuSerGlnGlnGlnLcLcLcLcLcLcLcLcLcLcLcLcLcLc 167
|||||
697 CTTCTTCACCACTTCATTCCTCAACCAAAATTCACCAAGTTC 746
|||||
167 GcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLc 184
|||||
747 GAGGCAAGGCTGAGTCAAGCAATCCTGATGCTGCTGCTGCTG 796
|||||
184 ValAlaSerProGlyThrSerGlnLcLcLcLcLcLcLcLcLcLc 200
|||||
797 CAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 846
|||||
201 GcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLc 217
  
```

```

|||||
847 GAGGCAAGGCTGAGTCAAGCAATCCTGATGCTGCTGCTGCTG 896
|||||
217 ValAlaSerProGlyThrSerGlnLcLcLcLcLcLcLcLcLcLc 234
|||||
897 TCGGCTGAGGCTAGCTGCAAGAGTCTCTGCTGCTGCTGCTG 946
|||||
234 alThrProLysGlnLysAlaAspProLcLcLcLcLcLcLcLcLc 250
|||||
947 TACGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 996
|||||
251 GlnArgPheAlaHisPheThrGlnLcLcLcLcLcLcLcLcLcLc 267
|||||
997 CAACCTTCTGCACTTCACCTGAGCTGAGTATCTGATCTGCAAG 1046
|||||
267 GValAspPheAlaLysGlnValProGlyPheLcLcLcLcLcLcLc 284
|||||
1047 GcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLc 1096
|||||
284 spcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLc 300
|||||
1097 ACAGATTCGCTCTCTGAGAGCTGAGTCAATGATGATGCTG 1146
|||||
401 ThrAlaArgArgTyrAsnHisSerThrGlnGlySerLcLcLcLc 417
|||||
1147 AGGCTCAAGCTTACCAACCAAGCAATCAATATATGCTTCTCA 1196
|||||
417 PProThrTyrSerLysAspPheProLcLcLcLcLcLcLcLcLcLc 434
|||||
1197 CTTCACCTACCAAGAGGCTGCTCTGAGCTGCTGCTGCTGCTG 1246
|||||
434 IleLeuAsnProLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLc 450
|||||
1247 TCATCAATTCATCTTCTGAGTCTCTGAGTATGCTGCTGCTG 1296
|||||
451 AspAspAlaGlnLysTyrAlaLcLcLcLcLcLcLcLcLcLcLc 467
|||||
1297 GAGATGATGAGGATGCTTCTGCTGCTGCTGCTGCTGCTGCTG 1346
|||||
367 PArgProAsnValGlnLcLcLcLcLcLcLcLcLcLcLcLcLcLc 384
|||||
1347 GCGGCTCAAGTGTAGAGAGGCTGAGGCTGCTGCTGCTGCTG 1396
|||||
384 YValGlnAlaLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLc 400
|||||
1397 ATGTGCAAGCTTCTCTCTCTCAACCAATCAATCAATCAATCA 1446
|||||
401 LeuArgPheProArgMetLysMetLysLysValSerLeuArgThr 417
|||||
1447 CTGCGCTTCCCAAGATGCTGATCAAGATGCTGATGCTGCTG 1496
|||||
417 PProValHisSerGlnGlnValPheAlaLeuAlaGlnAspLys 434
|||||
1497 GCGGCTCAAGTGTAGAGAGGCTGAGGCTGCTGCTGCTGCTG 1546
|||||
434 GcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLcLc 446
|||||
1547 TCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1594
|||||
sequence gb_F02.RN014533
  
```

seqname: gb_F02.RN014533
 seqdocumentation_block: 1959 bp. mRNA. 98-N-V-1-4
 DEFINITION: Ratius norvegicus orthologous receptor cDNA.
 ACCESSION: U14533
 VERSION: U14533.1 GI:565661
 KEYWORDS:

ORGANISM: Norway rat.
 Ratius norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Scurionathia; Muridae; Murinae;
 Ratius.
 REFERENCE: 1 (bases 1 to 1959)

|||||
 1497 CTGGTGACTGGAGTAGGTTTGTATGGCTTCAGGAAAGAAC 1546
 434 cnpProteinSerGluLeuPaspValHisGlu 446
 |||||
 1547 TGGCGCTTGGTGGTGGAGATCGATGTGATGATGAG 1564

seq.name: qb_pat1:AK045536

seq.document_id: block:

LOCUS AK045536 1979 bp DNA PAT 29-SEP-1999

DEFINITION Sequence 1 from patent US 5871916.

ACCESSION AK045536

VERSION AK045536.1 GI:5952204

KEYWORDS

SOURCE Unknown

ORGANISM Unknown

UNCLASSIFIED

REFERENCE 1 (bases 1 to 1979)

AUTHORS Nakamura, Y. and Saito, H.

TITLE cDNA protein and DNA encoding the same

JOURNAL Patent: US 5871916-A 1 16-FEB-1999

FEATURES Location/Annotations

SOURCE 1..1979

BASE COUNT 405 a 641 c 603 g 330 t

ORIGIN

alignment_scores:

Quality: 2033.50 Length: 491

Ratio: 4.751 gaps: 5

Percent Similarity: 92.842 Percent Identity: 87.202

alignment_block:

US-09-365-576-3 x AK045536

Align seq 1/1 to: AK045536 from: 1 to: 1979

1 MetSerSerPro...TherSerSerAspThrProValProGluValAsnGlu 16
 |||||
 206 ATGTCTCTCTTACACAGAGTTCACGACAAACCCCTGACGCAAAAGG 255
 15 YserProGluProSerThrSerAlaThrSerProThrLeuGluGluG 33
 |||||
 256 CCCCCCTCAACCTGCCCCCTTCTTCTTACCCACTGAAAGACAGAG 395
 33 LysGlnGlu.....ThrAspProPro...ProLysSerGlu 43
 |||||
 306 GTGGGAGGCGTGGCGGGGGGCGGAGGACATGTCGCAAGGACGAT 355
 44 GlySerSerSerAlaTyrLeuVal.....ValLeuGluProGluAs 58
 |||||
 356 GAGGTCAGATTCATCTGACAGACACAACTGGGTCATCTCATATCCAAAGA 405
 58 PGIuProGluAlaLysAlaLysLysLysLysLysLysLysLysLysLys 75
 |||||
 406 GGAACCTAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 455
 75 IsoluLeuGlyAlaValLysGlyLysPaspValAspSerLysPaspVal 91
 |||||
 456 ACGAGCTTGGCTGTCTGTGGGAGCAAGGCTCGCGCTTCGACATGACAG 505
 92 ValLeuSerSerSerLysLysLysLysLysLysLysLysLysLysLys 108
 |||||
 506 GTGTAACTGCAAGGCTGCAAGGCTTCTTCTGCGGCAAGGCTGCTGG 555
 108 SGIuAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 125
 |||||
 556 TGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 605
 125 SGIuAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 141
 |||||

606 ACGTTTCATGCGGAGCAAGTGGCAGTATGTCGTCGCAAGTCAGAG 655
 142 GluAlaGlyMetArgGluGlnGlyValLeuSerGluGluGluLeuArg 158
 |||||
 656 GAGGACAGGAGTAAAGGAGAGTGGCTCTCTCTGAGAGAGACATCCGAA 705
 158 LysArgLysGluGlnGluGlnGlu.....GluP 169
 |||||
 706 CAAGAGATTGGGAAGAGCAGCAGCAAGCAAGTCAAGTCACAGTGGTAT 755
 169 LysProProProSerGluProAlaAlaSerSerSerLysArgProAla 185
 |||||
 756 GATGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 805
 186 SerProGlyThrSerGluAlaSerSerGlnGlySerGlyGluGluGlu 202
 |||||
 806 TCCCTGTGTATGTAGAGTAGGCAAGGATGCTCTGAGGAGGAGGAGGAG 855
 202 YLeuGlnLeuThrAlaAlaGlnGluGluMetLeuGlnGlnLeuValAla 219
 |||||
 856 TGTCCAGCTAAAGAGGATCAAGAGCTAAATGATCAAGAGCTGTGGTGG 905
 219 LysLeuSerGlnPaspValLysArgSerPaspPaspPaspPaspValThr 235
 |||||
 906 CCAACAGAGAGAGTAAAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 955
 236 ProProProLeuAlaLysProGluThrArgPaspAlaArgGluGluArg 252
 |||||
 956 CCGCGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005
 252 GPhuAlaHisPheThrGluGluGluAlaLeuLeuValGlnGlnGlnGln 269
 |||||
 1006 CTGTCCCACTTCAAGAGATGAGGATGAGGATGAGGATGAGGATGAGG 1055
 269 SPhuAlaLysGlnValProGlyPheLeuThrGlyArgGluPaspGlu 285
 |||||
 1056 ACTGGCTAAGTAAATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1105
 286 LysAlaLeuLeuLysAlaSerThrLeuGluLeuMetLeuLeuGlnThrAla 302
 |||||
 1106 ATCCGCTCTGCAAGAGATCCAGATATGCAATATGCTCTGAGACAGAG 1155
 302 GArgArgTyrPaspHisGlnThrGluGlnGlnGlnGlnGlnGlnGlnGln 319
 |||||
 1156 CAGCGGCTACAACTACAGACAGAGAGGATGATATCTCTGAGAGCACTCA 1205
 319 ThrTyrSerLysPaspPheHisArgAlaLeuGluGlnValGluPheGlu 335
 |||||
 1206 CTAGACAGCAAGCAAGCACTTCACCGCTGAGGCTGAGGCTGAGGCT 1255
 336 AsnProLysPheGlnThrSerArgAlaMetArgArgGlnGlyLeuAspPasp 352
 |||||
 1256 AAGCCACTTTCAGATTCGCGGAGGATATGAGGCTGAGGCTGAGGCA 1305
 352 PAlaGluTyrAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 369
 |||||
 1306 GCTGAGTACGCTGTGCTATGCCCATCAATATCTCTGCGGCAATCGAG 1355
 369 LysValGlnGluProSerArgValGlnGlnAlaGlnGlnLeuProTyrVal 385
 |||||
 1356 GCAAGCTGCAAGCAAGCGGCGGCTGAGGAGGCTTCACAGAGCGCTAC 1405
 386 GluAlaLeuLeuSerTyrThrArgGlnGlnArgProGluAspGlnLeuArg 402
 |||||
 1406 GAGCGGCTGTGTGTATCAAGCGGCAATCAATATGAGGAGGAGGAGGAG 1455
 402 GPhuProArgMetLeuMetLysLeuValSerLysArgThrLeuSerSer 419
 |||||
 1456 CTGGGCGGAG 1505
 419 GHisSerGlnGlnValPheAlaLeuArgGluGlnAspLysLysLeuPro 435
 |||||
 1506 TGTACTGGAGTAAATCTTGGCTTGGGCTCTGAGGAGGAGGAGGAGGAG 1555


```

seq_name: gb-pat2:146765
seq_documentation_block:
LOCUS       146765               1898 bp    DNA                PAT          07-mar-1997
DEFINITION   Sequence 1 from patient US 563616.
ACCESSION    146765
VERSION      146765.1   GI:2470740
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unknown.
REFERENCE    1. (bases 1 to 1898)
AUTHORS      Liao,S., and Song,C.
TITLES       Isolated nucleic acid encoding a ubiquitons non heat resistant protein
JOURNAL      Patent: US 563616-A 1 17-JUN-1997;
FEATURES             location/qualifiers
source              1..1898
                     "for immunization known"
BASE COUNT        436 a      599 c      541 g      322 t
ORIGIN
alignment_scores:
                  quality: 2623.00           length: 461
                  ratio: 4.738              gaps: 5
Percent similarity: 92.625      Percent identity: 86.561
alignment_block:
US 09-365-576-3 x 146765 ..
Align seq 1/1 to: 146765 from: 1 to: 1898
1 MetSerProThrIleSerSetLeuAspThrProValProGlyAsnGlyse 17
|||||.....
71 ATGTCTACTCCCAACAAGTCCTCGTGATACCGCTGTGGCTGGAAATGGGC 120
|||||.....
121 CCTTAGCGCTGGCGCGCCTGGCTTTTCACCCACATCAAAGAAGAGGGCTC 170
|||||.....The Asprothro...ProLysSerGluLeu 44
34 InGln.....
|||||.....
171 CGAAGCGGTGGTGGCGGAGGAGTGGAGCGCATATGTCGACGAACATAATG 220
45 SerSerSerAlaTyrIleVal.....ValIleLeuGluProGluAspPst 59
|||||.....
221 GCCAGCTCAGCCTCCAGACACACATGGGTATGGTCAATGCCAATCCAAGAAC 270
59 uproGluArgLysArqLysGlySerGlyTrpAlaLysGlySerMetLysGlyHisG 76
|||||.....
271 ACCAAGAGCCACAGAAAAGAAAGGCGGCGCACGCGCGAAGCAATCGGGCGAC 320
76 IuhGcysArgValCysSolYaspLysAlasercLyphetStyArgSuval 92
|||||.....
321 AGCTTTGGGGTGTCTGTGGGACAAAGGCGCTGAGATTGCATTCAGATCAAGCTG 370
93 LeuSerGylGluGlyCysLyScyLythePharArgArgSerValValHist 109
|||||.....
371 CTGCAATTCAGAGACCTGCAACCGCTCTCTCCGCGCACAGTGTGTCGGTGT 420
109 pslYrAlaGlyArqTyrAlaGlyZarGlySerGlyTrpGlySerMetLysPsa 126
|||||.....
421 TGGAGCGCGAGGCGCTATGCTGTGCGGGGTGGCGGAAACCGCTGACATCGATG 470
126 IuThomcAtArgArqLysGlyScIuhGcysArqLysArqLysGlySerLys 142
|||||.....
471 CTTTAATGCGAGCGCAAGTGGCAAGTATGGTGGATTCGGCAAGGCGCAAGGAG 520
143 AlaGlyMetArgGluGlnGlyValIleSerGluGlnGlnIleValArgLysGly 159
|||||.....

```



```

151 AACCAAGACTGAGTCACTCCAAATTCGCAAGCAAGCAAGCAAGCAAGCAAG 200
64 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 81
201 AAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 250
81 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 97
251 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300
98 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 114
401 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 350
114 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 131
451 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 400
131 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 147
401 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 450
148 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 164
451 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 500
164 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 175
501 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 550
175 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 191
551 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 600
192 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 208
601 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 650
208 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 225
651 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 700
225 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 241
701 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 750
242 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 258
751 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 800
258 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 275
801 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 850
275 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 291
851 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
292 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 308
901 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 950
308 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 325
951 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1000
325 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 341
1001 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1050
342 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 358
1051 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1100

```

```

358 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 475
1101 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1150
375 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 491
1151 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1200
392 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 408
1201 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1250
408 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 425
1251 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1300
425 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1441
1301 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1450
442 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 446
1351 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1365
seq_name: qb_1h1q16:AC073806
seq_documentation_block:
ID: AC073806 215601 bp DNA HTG 18-JUN-2000
DEFINITION Mus musculus clone Rp23-47N6, WORKING DRAFT SEQUENCE, 22 ordered
pieces.
ACCESSION AC073806 GI:9256408
VERSION AC073806.2
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 215601)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 215601)
AUTHORS DOE Joint Genome Institute.
TITLE The Joint Genome Institute.
JOURNAL Submitted (29-JUN-2000) Product Gen Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
on Jul 18, 2000 (this sequence version replaced at:881042).
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1755238
Center clone name: Rp21-23_47N6
-----
Summary Statistics
Consensus quality: 203615 bases at least Q40
Consensus quality: 211090 bases at least Q30
Consensus quality: 212645 bases at least Q20
Estimated insert size: 220000; average-IP estimation
Estimated insert size: 214601; sum-of-counts estimation
Quality coverage: 7.88 in Q20 bases; average-IP estimation
Quality coverage: 8.1 in Q20 bases; sum-of-counts estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and

```



```

112862 CGAAGCAAGTACGCAAGGCTTCAGAGGAAAGGAGGAGATGCAATGAC 112911
207 AATAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 223
112912 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 112961
223 SASNGSAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 235
112962 CAAAGGAGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 113011
235 .....
113012 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113061
236 .....
113062 TGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 113111
239 .....
113112 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113161
239 .....
113162 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113211
239 .....
113212 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113261
239 .....
113262 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113311
239 .....
113312 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113361
239 .....
113362 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113411
239 .....
113412 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113461
239 .....
113462 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113511
239 .....
113512 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113561
239 .....
113562 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113611
239 .....
113612 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113661
239 .....
113662 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113711
239 .....
113712 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113761
239 .....
113762 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 113811
239 .....

```

```

315 GULYSASPHERETRYSETYSPASPHERHSAVALACTY...LEU...430
113812 TAAAGGATTTCAATACAGCAAGGAGATTTCAATACAGGATTTCAAG 113861
330 .....
113862 AAGGAGTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113911
330 .....
113912 GAGGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 113961
330 .....
113962 GAGGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 114011
330 .....
114012 AAGGAGTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 114061
331 .....
114062 TAAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 114111
336 .....
114112 TCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 114161
353 .....
114162 TAAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 114211
370 .....
114212 AAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 114261
386 .....
114262 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 114311
397 .....
114312 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 114361
397 .....
114362 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 114411
397 .....
114412 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 114461
398 .....
114462 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 114511
414 .....
114512 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 114561
430 .....
114562 GAGGCTGAGAGTGTATGATCTAGTACGATGAGTTCGCGAGAGTGGAG 114610
seq_name: gb_1:MM0132602
seq_documentation_block:
LOCUS MM0132602 6865 bp DNA 04-MAR-2000
DEFINITION Mus musculus LXR-beta gene, exons 1-10.
ACCESSION AJ132602
VERSION AJ132602.1 GI:5689852
KEYWORDS LXR beta gene; nuclear oxysterol receptor LXR-beta.

```



```

|||||
4533 ATGTCTCTGGCCAAAGATCTCTGAACTCTGCTGGCTGGTGAATACG 4582
12 .....
4584 GAGCTTCTTCAACCAACCACTCTCAATAGCTT GAGAACCTCACTGCTCT 4632
13 .....Pro-GlyAspGlySerProGlnProSerHisSerAlaIle 25
4633 GTCTTCTTCTTCTTATAGCAATGGTTCTCTCAATCCAGTACCTCTGGTAC 4682
25 rSerProHisIleGlySerIleGlySerIleHisAspProProGlyLeu 42
4684 GTCATCTCACTATTAAGCAAGAGGGGGTAAAGAACCTGATCTCTCTCAAGCT 4732
42 GTCTGCTGlySerSerSerAlaIleTyrTle..... 50
4733 CTGAAGAGCTCAAGCTCTCAATGATGGTGGATGGAGATGGGGAGCAAGGAA 4782
50 .....
4784 CAAGGCTCTGAGAGCTTCTTGAGAGCAAGGGGGTGGAGATGGTGGTCTGT 4832
50 .....
4833 CTCTTGGTGGAGTCTTGGCTTTCCTTTCCTGCTGAGCAAGAGCACTGCTCT 4882
50 .....
4884 CGGAGAGCTGGCAACCACTCAATCACTTTCCTTTCGGCTTGGAGCCCTGGC 4932
50 .....
4933 CTCTAGCTCTGCTTTTGGAGCTCTGCAATGGCTTTCCTGAGAGCTGCTCTT 4982
51 .....Val...ValIleLeuG 55
4983 CTCTTCTGCTCAATCTCTCAATGGCAATCTGCTTACCCCAAGCTCAATGAGA 4032
55 GProGluAspGlnProGlnAlaIleGlySerIleGlySerIleProAlaIleGlySer 72
4033 GGTAAAGAGATCAATCTCAAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 4082
72 CTCTGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 88
4083 TCTCTGAGCTCAATCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4132
89 HisIlePheAspValIleGlnSerIleGlySerIleGlySerIleGlySerIle 105
4133 CACTCAATCAAGCTCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4182
105 PAlaValIleHisGlySerIleGlySerIleGlySerIleGlySerIleGlySer 122
4183 TCTGCTCTCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4232
122 ySerIleMetAspAlaIlePheMetArgAlaGlyGlySerIleLeuGlySerGln 138
4233 GTCATATGATATCTTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4282
139 HisGlySerIleGlySerIleGlySerIleGlySerIleGlySerIleGlySer 150
4283 AAGTCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4332
150 .....
4333 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4382
151 .....LeuSerIleGlnGlnIleLeu 157
4383 GAGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4432
157 GlysIleSerAlaIleGlnIleGlySerIleGlnIleGlnIleProProProSer 174
|||||

```

```

4433 GAAGAAAGAGATTCAAGAACCACTCAAGCAGCAATCAACCAACCAATCTG 4482
174 LProAlaIleAspSerSerGlyAlaIleProAlaIleSerProGlyIleHis 190
4483 AGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4532
191 GluAlaSerSerGlnIleGlySerIleGlySerIleGlySerIleGlySerIle 207
4533 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4582
207 AlaGlnGlnIleLeuMetIleGlnIleLeuValIleAlaIleLeuHisGly 224
4583 GGTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4632
224 SolIleArgSerPheSerAspIleProIleValIlePhe..... 235
4633 ACNAAAGATCTTCTGCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4682
235 .....
4683 GCTCTGTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4732
236 .....ProIleProLeuG 240
4733 CGAGTCTCCCAAGCTGAGCCCTGGCTTCTTGTGTGGAGAGAGAGAGAGAG 4782
240 LValAspProGlnSerAlaAspAlaIleArgIleGlnArgPheAlaHisPhe 256
4783 GTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4832
257 ThrGlnLeuAlaIleIleSerValIleGlnIleIleValAspPheAlaIleGly 273
4833 ACAGAGCTAGGTATCATCTGGGTCTCAAGAGATGTGGATTTGGTAAAGCA 4882
273 HisAlaProGlyPheLeuGlnHisGlyArgGlnAspIleIleAlaIleLeu 290
4883 GGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4932
290 ySAlaSerThrIle..... 294
4933 AGGCTGCAATCTTGAAGATATCAATATATAGAGTATAGAGAGAGAGAGAG 4982
294 .....
4983 GAACATGATAGAGCAGATGCTGAGGCTTGGGGTGTATGATGATGAGAGAG 5032
294 .....
5033 GGTCTGATGCAAGAGATCTGAGAGCTCTAGAGAGAGAGAGAGAGAGAGAG 5082
294 .....
5083 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5132
294 .....
5133 TTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5182
294 .....
5183 AATCAAGAGATATGAGATATGATAGATTAACCTGCTGAGAGAGAGAGAGAG 5232
294 .....
5233 GGTAAATATTAAGCTTACGCTGAGGTGAGAGATGATGATGATGATGATGAT 5282
294 .....
5283 CCTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5332
294 .....
5333 GTCTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5382

```


cc plasmid pGEN-2, as an EcoRI fragment, to form pR8-15p6. The
cc encoded protein is involved in modulation of telomeric acid
cc signaling.

xx Sequence 1944 bp: 440 A: 588 G: 544 G: 362 T: 0 other:

alignment_scores:

quality: 224.00 length: 416
Ratio: 5.121 gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.286

alignment_block:

us-09-365-576-3 x T13229

Align seq 1/1 to: 113229 from: 1 to: 1934

```

1 MetSerSerProThrSerSerLeuAspThrProValProGlyAsnLysLe 17
|||||
191 ATGCTTCTCCCAATAGCTCTGGACATGCTCTGCTGGTGGTAAATGGTTG 240
17 ProGlnProSerThrSerAlaThrSerProThrLLeLysGlnGlyG 34
|||||
241 TCCGCAAGCCAGTACTCTCTCCACTTCACCCATATTAAAGAAAGAGAGC 290
34 LngLutAspProProGlySerLysSerLysSerSerAlaLysLle 50
|||||
291 AGCAATCTGATCCATCCCAAGCTCTGCAAGCCCTCCCACTCTCTTACATC 400
51 ValValLLeLeuGlnThrProLAspGlnProGlnThrGlySerL 67
|||||
341 GTGCTGATCTTAAAGTAAAGAGATGAAGCTGAAGCGAAGAGAAAGAG 490
67 yProAlaProLysMetLeuGlnLysLysLysLysLysLysLysLys 84
|||||
391 TCCGCGCCGCAAGATCTCTGAGCATGATGTCGGCGCTGCGGAGACA 440
84 ySAyLysSerLysProLysLysLysLysLysLysLysLysLysLys 100
|||||
441 AGGATCTCCGCTCTGCAATCAATGCTCTGCTGCTGCTGCTGCTGCTG 490
101 PhePheAspAspSerValValLLeLysGlnGlyAlaGlyAlaLys 117
|||||
491 TTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
117 GcLysSerGlyThrLysLysLysLysLysLysLysLysLysLys 144
|||||
541 GAGCAAGAGCAAGCTCTGCAATGATGCTCTGCTGCTGCTGCTGCTG 590
134 GcLysAlaGlyAlaLysLysLysLysLysLysLysLysLysLys 150
|||||
591 TCTGATACATGCTGCAATGCAAGAGCTGCTGCAATGCTGCTGCTGCT 640
151 LeuSerGlnGlnGlnLLeLysLysLysLysLysLysLysLysLys 167
|||||
641 CTCTCTGAGGAGGAGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 690
167 GcLysAlaGlyAlaLysLysLysLysLysLysLysLysLysLys 184
|||||
691 GCAATCAATGCTGCAATGCAAGAGCTGCTGCAATGCTGCTGCTGCT 740
194 LLeLysSerProGlnThrSerGlnAlaLysSerSerLysLysLys 209
|||||
741 CAGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 790
201 GcLysLLeLysLysLysLysLysLysLysLysLysLysLysLys 217
|||||
791 GAGGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
217 LLeLysLysLysLysLysLysLysLysLysLysLysLysLysLys 244
|||||
841 TCCGCGCCGCAAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 890

```

```

244 ATThrProTProLLeuGlnValAspProGlnLLeuSerAlaAspAlaLysGln 250
|||||
891 TCCGCGCCGCAAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 940
251 GLeuGlnPheAlaLysSerProThrGlnLysLysLysLysLysLysLys 267
|||||
941 CAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 990
267 ValAspPheAlaLysSerLysValLLeuGlnGlnGlnGlnGlnGlnG 284
|||||
991 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1040
284 spGlnLLeuAlaLysLysLysLysLysLysLysLysLysLysLysL 400
|||||
1041 AGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1090
301 ThrAlaAspAlaGlyAlaAsnLysGlnLLeuGlnLLeuGlnLys 417
|||||
1091 AGCAGAGAGAGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
317 PheThrTyrSerLysAspAspPheAspAlaLysLysLysLysLysL 344
|||||
1141 CTTCAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1190
344 LLeuAspProLLeuGlnLLeuPheSerAlaLysLysLysLysLysL 450
|||||
1191 TCAATCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1240
351 AspAspAlaGlyAlaLysLysLysLysLysLysLysLysLysLys 467
|||||
1241 GAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1290
467 PheProAspValLLeuGlnLLeuProSerAlaLysLysLysLysL 384
|||||
1291 TCCGCGCCGCAAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1340
1291 CCGCGCTGCAATGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1340
384 yValGlnAlaLysLysLysLysLysLysLysLysLysLysLysL 400
|||||
1341 ATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1390
401 LeuAspPheProAlaGlnLLeuMetLysLysValSerLysLysL 417
|||||
1391 CTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
417 LysValLysSerGlnGlnValPheAlaLysLysLysLysLysLys 444
|||||
1441 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1490
444 GcLysProLLeuLysSerGlnLLeuTyrAspValLysLysL 446
|||||
1491 TCCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1528

```

seq_name: us-09-365-576-3 from: 1 to: 1934 us-09-365-576-3

seq_documentation_block:

ID: Q88761 standard: DNA: 1959 bp.

AC: Q88761:

DT: 20-JAN-1996 (first entry)

EE: Rat ubiquitous nuclear receptor protein RNA.

KW: Ubiquitous nuclear receptor drug binding; disease diagnosis;

KW: therapy; ss.

OS: Rattus rattus.

PH: Key location/Qualifiers

FT: CDS 256..1919

PN: W09513473-A1.

```

XX 18 MAY-1995.
XX
XX 08 NOV-1994: 94MO-US1984
XX
XX 10 NOV-1994: 94OS-0152004
XX
XX (ARCH.) ARCH. DEV. CORP.
XX
XX Liao S., Song C.
XX
XX WP1: 1995, 194092/25.
XX P-PSDB: R74749.
XX
XX New ubiquitous receptor used to develop precls. for use in
XX diagnosis, drug design and therapeutic applications.
XX
XX Claim 6: Page 151, 199pp: English.
XX
XX This genomic DNA isolated from rat brain may be cloned
XX recombinantly to produce a cDNA which interacts with the response
XX elements and network of receptors in the thyroid hormone/prolifer-
XX ation receptor substantially. The products can be used in detection,
XX diagnosis and in screening assays for substances which interact
XX with cDNAs for use in diagnosis, drug design and therapeutic
XX applications.
XX
XX Sequence 1959 bp: 486 A, 607 C, 565 G, 381 T, 0 other.

alignment_scores:
    quality: 2247.40    length: 446
    ratio: 5.085        gaps: 1
    percent similarity: 99.104    percent identity: 97.085

alignment_block:
    us-09-365-576-3 x Q88761
    Align seq 1/1 to: Q88761 from: 1 to: 1959

1 MetSerSerProThrSerLeuAspThrProValProGlyAsnGlySer 17
  |||
256 ATGCTCTTCCCAAGATGCTCTGAACTGCTCTCGGAATGGTGC 305
  |||
17 rProAlaProSerThrSerAlaThrSerProThrTleGlyGluGly 34
  |||
406 TCTGCAAGCTCAGTACGCTGCTCACTTCAACCACTAATAGGAGGAGTAC 355
  |||
44 InclThrAspProProGlySerGlyGlySerSerAlaTyrIle 50
  |||
456 AATGAGATGATGACCTGCTGAACTGCAAGCTGCTGCTTACATC 405
  |||
51 ValValIleLeuAlaProGlyAsnGlyProGlyAlaValGlySer 67
  |||
406 GTC.....GAGCTAGAGATGAACTTCAAGCAAGGAGAAAGAG 446
  |||
67 yProAlaProGlyMetIleGlyIleGlySerGlyAspValGlySer 84
  |||
447 TCTGAGGAGGAGATGCTGAGGAGGAGGAGGAGGAGGAGGAG 496
  |||
84 ysaIAspSerGlyPhenGlySerValLeuGlySerGlyGlyGly 100
  |||
497 AATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 546
  |||
101 PhenAlaGATGATSerValValIleGlyGlyAlaGlyAlaGlySer 117
  |||
547 TCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 596
  |||
117 yglIysrGlyThrGlySerMetAspAlaPhenMetArgAlaGlyGly 134
  |||
597 GGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 646
  |||
134 GATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 150

```

```

|||||
647 TGTGAGAGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696
  |||
151 rousSerGlyGluGlnIleArgGlySerArgIleGlyGluGluGln 167
  |||
697 GTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 746
  |||
167 nGlnProProProSerGluProAlaAlaSerSerSerGlyAlaGln 184
  |||
747 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 796
  |||
184 AlaAspProGlyThrSerAlaAlaSerSerGlyGlySerGlyGly 200
  |||
797 GATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 846
  |||
201 GluGlyIleGlySerValAlaAlaGlyGluGlyMetIleGlnIle 217
  |||
847 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 896
  |||
217 AlaAlaGlnIleGlySerValAlaAlaGlyGlySerGlySerGly 234
  |||
897 TCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 946
  |||
234 AlrProIleProGlyAlaAlaAspProGlnSerArgAspAlaGln 250
  |||
947 TCAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 996
  |||
251 GlnArgPheAlaHisPheThrGluGluAlaIleIleSerValGln 267
  |||
997 GAGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1046
  |||
267 GValAspPheAlaGlySerValProGlyPheGlnGlnGlnGln 284
  |||
1047 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1096
  |||
284 spGlnIleAlaLeuLeuGlyAlaSerThrIleGlnIleMetLeuGln 300
  |||
1097 ACCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1146
  |||
301 ThrAlaArgArgIleAsnHisGlyThrGlyGlySerIlePheGln 317
  |||
1147 AATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1196
  |||
317 PheThrTyrSerGlyAspAspPheHisArgAlaGlyLeuGlnVal 334
  |||
1197 GTTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1246
  |||
334 IleIleAsnProIlePheGlyPheSerArgAlaMetArgArgLeuGln 350
  |||
1247 TATCAATGCCATCTTGAAGTTCTCTGAGGATATGAGGAGGAG 1296
  |||
351 AspAspAlaGlyAlaAlaLeuLeuIleAlaIleAsnIlePheSer 367
  |||
1297 GATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1346
  |||
367 PArgProAsnValGluGluProSerArgValIleAlaLeuGlnIle 384
  |||
1347 GCGAGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1396
  |||
384 yValGlnAlaLeuLeuSerTyrThrArgIlePheArgProGlnIle 400
  |||
1397 ATGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1446
  |||
401 LeuArgPheProArgGlnMetLeuMetLeuValSerLeuArgIle 417
  |||
1447 GTGAGCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1496
  |||
417 rSerValHisSerGlnGlnValPheAlaLeuArgLeuGlnAspGly 434
  |||
1497 GTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1546
  |||
434 rProProLeuLeuSerGlyIleTyrPheValHisGln 446
  |||

```


Quality: 2034.59 Length: 461
 Ratio: 4.751 Gaps: 5
 Percent Similarity: 92.842 Percent Identity: 87.202

alignment block:
 US-09-365-576-3 X 130031 ..

Align Seq 1/1 to: T30041 from: 1 to: 2030

```

1 MetSerPro...ThrSerLeuAspThrProValProGlyAsnGlu 16
  |||||
245 ATCTCTCTCTACGACGAGTTCCTGCGAACGGCTGCTGCGAATGCG 294
  |||||
16 ySerProGluProSerThrSerAlaThrSerProHu1HebGluGln 33
  |||||
295 TCCCTCTACGCTGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 344
  |||||
33 yGluGln...ThrAspProPro...ProGlySerGlu 43
  |||||
445 GTTCCGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 394
  |||||
44 GlySerSerSerAlaTyrTLeVal...ValTLeuGluProGluAs 58
  |||||
495 GAGCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 444
  |||||
58 pGluProGluProGlySerAlaTyrSerGlyProAlaTyrGlySer 75
  |||||
445 GAGCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 494
  |||||
75 pGluProGlySerAlaTyrSerGlyProAlaTyrGlySerGlu 91
  |||||
495 GAGCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 544
  |||||
92 ValLeuSerGlySerGluGlySerGlyProGluProGluValVal 108
  |||||
545 GTGCTGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 594
  |||||
108 pGlySerTyrTLeValTyrSerAlaTyrSerGlyProGlySerGlu 125
  |||||
595 GTGCTGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 644
  |||||
125 pGluProGlySerAlaTyrSerGlyProGluProGluValVal 141
  |||||
645 GTGCTGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 694
  |||||
142 GluAlaGluProGlySerGlyProGlySerGlyProGlySerGlu 158
  |||||
695 GTGCTGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 744
  |||||
158 pGluProGlySerGlyProGlySerGlyProGlySerGlu 169
  |||||
745 GAGCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 794
  |||||
169 pGluProGlySerGlyProGlySerGlyProGlySerGlu 185
  |||||
795 GAGCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 844
  |||||
186 SerProGlyThrSerGlyAlaTyrSerGlyProGlySerGlySer 202
  |||||
845 GTGCTGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 894
  |||||
202 yGluGlnGluThrAlaGluGluGluGluGluGluGluGluGluGlu 219
  |||||
895 GTGCTGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 944
  |||||
219 pGluProGlySerGlyProGlySerGlyProGlySerGlu 235
  |||||
945 GTGCTGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 994
  |||||
236 pGluProGlySerGlyProGlySerGlyProGlySerGlu 252
  |||||
995 GTGCTGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1044
  |||||

```

```

252 pGluAlaHisPheThrGluAlaLeuTyrSerValGluGluTyrVal 269
  |||||
1045 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1094
  |||||
269 pGluAlaHisPheThrGluAlaLeuTyrSerValGluGluTyrVal 285
  |||||
1095 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1144
  |||||
286 pGluAlaHisPheThrGluAlaLeuTyrSerValGluGluTyrVal 302
  |||||
1145 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1194
  |||||
402 pGluProGlySerGlyProGlySerGlyProGlySerGlu 319
  |||||
1195 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1244
  |||||
319 pGluProGlySerGlyProGlySerGlyProGlySerGlu 335
  |||||
1245 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1294
  |||||
336 pGluProGlySerGlyProGlySerGlyProGlySerGlu 352
  |||||
1295 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1344
  |||||
452 pGluProGlySerGlyProGlySerGlyProGlySerGlu 369
  |||||
1345 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1394
  |||||
369 pGluProGlySerGlyProGlySerGlyProGlySerGlu 385
  |||||
1395 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1444
  |||||
386 pGluProGlySerGlyProGlySerGlyProGlySerGlu 402
  |||||
1445 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1494
  |||||
402 pGluProGlySerGlyProGlySerGlyProGlySerGlu 419
  |||||
1495 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1544
  |||||
419 pGluProGlySerGlyProGlySerGlyProGlySerGlu 435
  |||||
1545 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1594
  |||||
436 pGluProGlySerGlyProGlySerGlyProGlySerGlu 446
  |||||
1595 GTTCCGACGACGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1644
  |||||
seq name: pGluProGlySerGlyProGlySerGlyProGlySerGlu
seq documentation block:
ID: T27616 standard, cDNA to mRNA, 1579 bp
XX
AC T27616
XX
DI 09-NA-1996 (first entry)
XX
DE Human foetal lung steroid hormone receptor analogue ECDN cDNA.
XX
KW Human; foetal lung; steroid hormone; receptor; analogue; protein;
KW ECDN protein; cDNA; sequence; binding domain; recombinant;
KW foetal lung; analogue; drug; various tissue; primer;
KW probe; antibody; immunohistochemical assay; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS
FT /tag= d
XX
PN W0900924-A1.
XX
PD 28-MAR-1996.

```



```

XX      ECDN protein, a steroid hormone receptor analogue from human foetal
PT      lung - is expressed in cancer cells and is useful for cancer
PT      diagnos and drug development
XX
PS      Claim 4: pages 26-28; 44pp; Japanese.
XX
CC      The present sequence encodes the variant of the human foetal lung
CC      derived steroid hormone receptor analogue protein ECDN, distinguished
CC      ECDN small mol. (ECDNsm) protein. ECDNsm protein is expressed in
CC      various cancer cells, therefore screening for ECDNsm protein
CC      binding moles., using recombinant ECDNsm proteins will be useful in
CC      the identification of candidate anticancer drugs. Gene expression
CC      of ECDNsm proteins in various tissues can be studied using
CC      primers and probes derived from ECDNsm protein. ChA. Acc. No. J1-10-1
CC      which recognises ECDNsm proteins can be used in ECDNsm protein
CC      immunohistochemical assays.
XX
80      Sequence: 1688 bp; 447 A; 562 G; 494 C; 285 T; 0 other.

alignment_scores:
      quality: 1480.00      length: 454
      ratio: 4.879      gaps: 6
      percent similarity: 74.614      percent identity: 66.674

alignment_block:
us-09-365-576-3 x 127617 ..
Align seq 1/1 to: 127617 from: 1 to: 1688

1 MetSerSerPro...ThiSerSerIuaspThiProValProValProValAsn 16
|||||
206 ATGTCCTCTCTACACACACACACACACACACACACACACACACACACACAC 255
16 ySerProGluProSerThiSerAlaThrSerProThiAlaGlyGluGlu 43
|||||
256 CCGCCCTGACCTGGGCGCCCTCTCTCTGACGACCTGTAAGAGGAGAG 305
43 yThrGluThrAspProGluProGluGlySerGlyGlySerGlySerAla 49
|||||
406 GTGGGAGG.....CGTGGGCGGAGG..... 425
50 IleValValIleLeuGluGluProGluIuaspGluProIuAlaArgGlyGly 66
|||||
425 ..... 425
66 sGlyProAlaProGlySerGluGluGlyIleGluLeuGlySerAlaGlyGly 83
|||||
426 GTTCGAGACCTGATGTCGACGACCTAT..... 255
83 sPylsAlaSerGlyPylsHisIlyAsnValLeuSerGlySerGlyGlyGly 99
|||||
355 ..... 355
100 GlyIlePheAlaPheSerValValHisGlyGlyAlaGlyAlaGlyAlaGly 116
|||||
355 ..... 355
116 sAlaGlySerGlyThrGlySerIuMetAspAlaPheMetAlaIleGlyGlySer 133
|||||
356 .GAGGCGACGCTGACGCTGACGACACACAC..... 382
143 IuMetGlySerAlaGluAlaGlySerGlyGlyAlaGlyAlaGlyAlaGlyIle 149
|||||
383 ..... 383
150 ValLeuSerGluGluGluIleAlaGlySerAlaGlyGlyGlyGlyGlyGly 166
|||||
489 GTCTCTCTGACACACACACACACACACACACACACACACACACACACAC 438
166 rGlu.....GluProProProProSerGluProAla 177
|||||

```

```

439 GAGAGATTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 488
177 IAsrSerSerGlyAlaProAlaAlaSerProGlyThiSerGluAlaSer 194
|||||
489 GACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 538
194 SerGluGlySerGlyGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 210
|||||
539 ACCAGACCTGTCGAGACCTGTCGAGACCTGTCGAGACCTGTCGAGACCTG 588
210 sPylMetIleGluIleValAlaValAlaGluGluGlySerGlySerGlySer 227
|||||
589 ACTATATATTCACAGTATGTCGAGGAGGTCGAGTATGTCGAGTATACAAAG 638
227 cGluMetAspGluThrGlyValIlePheIleGluGluGlyAlaGlyPro 243
|||||
639 CATTGCTGACACACACACACACACACACACACACACACACACACACAC 688
244 GluSerAlaAspAlaArgGluGluAlaGluAlaHisPheThrGluAla 260
|||||
689 CATTGCTGACACACACACACACACACACACACACACACACACACACAC 738
260 aIleIleSerValGluGluIleValAspPheAlaGlySerIleValProGlyP 277
|||||
739 CATCATCTCATGTCGAGACCTGTCGAGTATGTCGAGTATGTCGAGTATG 788
277 bGluGluIleGluGlyAlaGluAlaSerGluAlaGluAlaGluAlaSerThr 294
|||||
789 TCTGACGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 838
294 ILeuIleIleMetLeuLeuGluGluIlePheAlaAlaGlyAlaGluGluGlu 310
|||||
839 ATGAGACATATGTCGAGACACACACACACACACACACACACACACACAC 888
310 cGlySerIleThrPheLeuGlyAspPheThrGlySerGlySerAspPheHis 327
|||||
889 GTGATATGCTCTCTGAGGAGCTTCACCTACATAGAGAGGAGCTTCGACG 938
327 rAlaGluGlyGluGluValGluPheIleAspProIleThrGluPheSerArg 344
|||||
939 GTGAGGAGGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 988
344 AlMetAlaPheAlaGluGlyLeuAspAspAlaGluGlyAlaGluAla 460
|||||
989 GCGTGTGCGGAGCTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1038
1039 CATGAGATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1088
377 aGluAlaGluGluGluGluProGlyValGluAlaGluGluGluSerGlyThrArg 994
|||||
1089 TCGAGGCTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1138
394 ILeuSerProGluIuaspGluLeuAlaGluGlyAlaGluMetLeuGlySer 410
|||||
1139 ATCAAGAGAGGCGGACGACGACGCTGCTCTCTCTCTCTCTCTCTCTCT 1188
410 uValSerIleArgThrLeuSerSerValHisSerGluGluValPheAla 427
|||||
1189 GGTGAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1238
427 cGluAlaGluGluGlyGlyGlyGluGluProGluGluGluGluGluGluGlu 444
|||||
1239 TCGGCTGTCAGGAGAGAGAGGCTGAGGCTGTCGTCGACATCTGAGGAG 1288
444 ValHisGlu 446
|||||
1289 GTCCACGAG 1297
seq_name: us-09-365-576-3 from: 1 to: 1688 bp; 447 A; 562 G; 494 C; 285 T; 0 other
seq_documentation_block:

```

10 Z45043 standard: cDNA: 1528 bp.
 XX
 AC Z45043;
 XX
 28 FEB-2000 (first entry)
 XX
 DE Human nuclear orphan receptor LXR-alpha cDNA.
 XX
 NM Nuclear orphan receptor LXR-alpha; human; rat; CNRBB-1;
 NM CNRBB-1; rat; rat; rat; rat; rat; rat; rat; rat; rat;
 NM cardiovascular disease; hypertension; congestive heart failure;
 NM diagnosis; therapy; transcription factor; ss.
 XX
 OS Homo sapiens.
 XX
 DN W095544.1.
 XX
 04-NOV-1999.
 XX
 23-APR-1999; 9900 DBR502.
 XX
 24-APR-1998; 9800 C082997.
 XX
 (B30M) BRIGHAM & WOMEN'S HOSPITAL INC.
 XX
 Chen YE, Horuchi M, Dzan VJ, Tanura K;
 XX
 WP1: 2000-062002709.
 XX
 A new regulatory element binding factor used to treat an adverse
 PF cardiovascular condition.
 XX
 claim 26; Page 77; 57pp; English.
 XX
 This is the nucleotide sequence of human nuclear orphan receptor
 XX LXR-alpha cDNA. It shows 898 homology to newly isolated cDNA (see
 XX Z45043) encoding mouse negative regulatory element binding factor
 XX CNRBB-1 (see Y42474). CNRBB-1 binds to the 5'-flanking region of
 XX the renin gene and modulates expression of the gene. CNRBB-1 related
 XX polynucleotides, polypeptides and inhibitors are used in claimed
 XX methods for increasing or decreasing CNRBB-1 activity in a
 XX mammalian cell or subject, e.g., for treatment of an adverse
 XX cardiovascular condition, hypertension, or congestive heart failure
 XX (claimed). In a claimed method, the CNRBB-1 inhibitor is an
 XX antisense nucleic acid or a dominant negative nucleic acid of the
 XX present sequence. A claimed method for determining the level of
 XX CNRBB-1 expression also utilizes the present sequence; an increase
 XX in CNRBB-1 expression being indicative of an increased
 XX susceptibility of a subject for developing a renin-angiotensin
 XX system mediated disorder such as hypertension or a cardiovascular
 XX disorder. LXR-alpha is a novel cAMP responsive transcriptional
 XX factor for human renin gene expression.
 XX
 580 Sequence 1528 bp; 440 A; 454 C; 439 G; 295 T; 0 other;
 alignment_scores:
 Quality: 1395.60 Length: 453
 Ratio: 4.844 Gaps: 6
 Percent Similarity: 89.142 Percent Identity: 62.472
 alignment_block:
 US-09-365-576-3 X Z45043
 Align seq 1/1 to: Z45043 from: 1 to: 1528
 4 prothrserrseuasprrhrprvaltrpolly...asnghysrrprogl 19
 75 cctgaattctgcctgagctctgtaaaccttaccggtacgatacag 124
 19 nprosrthrrseuatahrsrprothrllcysgclucjucjucjucjuc 36
 125 ccaaaacccagagacacacacacacacacacacacacacacacac 174

36 hrAspProthrProthrlyserGluThySerSerSer 47
 175 CCCACTCTGCTGGGGGTACTGTACAGGGGTGGAGGCTGGAGGCTG 224
 48 AlatyrlleValVallleuenglnProGlnAspGluPro..... 60
 225 ACAGCCCTGCTGTACAGGGGTACAGGGGTGGAGGCTGGAGGCTG 274
 61CAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 77
 275 TCCACAAAAGCCGAAAAAGGGGTACAGGGGTACAGGGGTACAGGG 324
 77 GCGSARVAVAGCSClyasrlyasrlyasrlyasrlyasrlyasrly 94
 325 TATGTAGGCTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCAAT 374
 94 SCGYSGluclycyslysclypPhaThrAatqsrValValnlsclly 110
 175 AGCTGTGAGTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCAAT 422
 423GCCACTACATCTGTGACAGTGGGGGTACAGGGGTACAGGG 468
 110 pAscllyAatqsrValValnlscllysclysclysclysclyscly 127
 127 hewetrAatqsrlysclysclysclysclysclysclysclyscly 143
 459 AATGGTGTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCAAT 518
 144 GtymcAagggcluclycysValleuenglnProGlnAspGluPro 160
 519 GGTATCGCGAGGAGAGTGTGTGGAGCAAGAGGCTGGGGTTCGAC 568
 160 gtleglnlysglnnglnnglnnglnnglnnglnnglnnglnngln 177
 559 ATGTGAAGGGCAAGAGGAGCAAGAGGCTGTGACATCTTGGGTTC 618
 177 lAsersersrcllyAatqsrValValnlscllysclysclyscly 194
 619 GGGTGTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCAAT 647
 194 srcllysclysclysclysclysclysclysclysclysclyscly 210
 648CAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 667
 219 GCGSARVAVAGCSClyasrlyasrlyasrlyasrlyasrlyasrly 227
 668 GGGATATGATGAGAGGCTGTGTGGAGCAAGAGGCTGGGGTTCG 717
 227 GtymcAagggcluclycysValleuenglnProGlnAspGluPro 243
 718 CTTTCTGTACGGGCTGTGTGGAGCAAGAGGCTGGGGTTCGACAT 767
 244 GlnserAatqsrValValnlscllysclysclysclysclyscly 260
 768 ATATGAGGGTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCA 817
 260 GtymcAagggcluclycysValleuenglnProGlnAspGluPro 277
 818 CATGTGTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCAAT 867
 277 hewetrAatqsrlysclysclysclysclysclysclysclyscly 293
 868 TGTGTGTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCAAT 917
 294 hewetrAatqsrlysclysclysclysclysclysclysclyscly 310
 918 ATGTGAAGGGTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCA 967
 310 nCysllcThrleuValAspPheThrlysclysclysclysclyscly 327
 968 GAGTATACACTTGTGTGTGGAGCAAGAGGCTGGGGTTCGACATCA 1017

• • • •


```

51 ValValIleLeuGluProGluAspLeuProGluArgArgAspGlySerG1 67
   |||
281 GAGGATATTAGAACTAAATATAGCTGAGGCAAGGGAAGAAAGG 330
   |||
67 yPpEALProGlyMetLeuGlyHisGluLeuGlyAspValGlySerG1 84
   |||
441 GCGAGGCGCGGAGAGATCGAGGAGATGCTGCTGGGCGTGGTGGGAGCA 380
   |||
84 ySALAspSerGlyPheHisGlyArgAspValLeuSerGlyGluGlySerG1 100
   |||
481 AGGATTGGGAGCTTCTAACTAACTGTGTAACTGTGAAGGCTTCAAGAGC 430
   |||
101 PhePheAspArgSerValValHisGlyGlyValArgArgPheArgGlySer 117
   |||
441 TCTTTCGGAGCTAGTGGTGTCTACGAGGAGCGGAGGCTTGGCTTGG 480
   |||
117 yGlySerGlyThrGlySerGluMetAspAlaPheMetArgArgGlySerG1 134
   |||
481 GAGGCAAGGAGAGCTGGTAAATGAGATGCTTCAATCGAGCGCAAGTGGCAGC 530
   |||
134 yGlySerArgGlyAspArgGlySerGlyGluGlyMetArgGluGlySerVal 150
   |||
541 TCTGGAGGCTGGGTAAGTGTAAAGAAAGCTGGTATGGGAGAGCTGGCTG 580
   |||
151 LeuSerGluGluGluGluGluGluGlySerArgGluGluGlySerGluGlu 167
   |||
581 GCTGCTGAGAGAGACATTCGAGAAAGAGATTCAGAGAGAGAGAGAGAGCA 630
   |||
167 MetLeuProProProProSerGluProAlaAlaSerSerSerGlyArgPro 184
   |||
631 GCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 680
   |||
184 LALISerProGlyThrSerGluAlaSerSerGluGlySerGluGlySerG1 200
   |||
681 TGGGCTGGTGGGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 740
   |||
201 GluGlyLeuLeuGluArgPheAlaAlaSerSerGluGluGluGluGlu 217
   |||
741 GAGGAGATTCAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
   |||
217 LALALALALALALALALALALALALALALALALALALALALALALAL 234
   |||
781 TGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 830
   |||
234 ALThrProProProProGluAlaAspProGluSerArgAspAlaArgGlu 250
   |||
831 TCGAGGCTGGGCGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 880
   |||
251 GluAspPheAlaHisPhePheHisGluAlaGluGluSerValGluGlu 267
   |||
881 TAACTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 930
   |||
267 yVALISerProGluAlaSerGluValProGlyPheGluGluGluGluGlu 284
   |||
931 TGTGGAGCTTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 980
   |||
284 spEALLeuAlaLeuLeuGlyAlaSerThrIleGluIleMetLeuLeuGlu 300
   |||
981 AGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1040
   |||
301 ThrAlaArgArgArgArgAspHisGluArgGlyIleThrPheLeuGlySer 317
   |||
1041 ACATCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 1080
   |||
317 GpPheThrTySerGlySerAspAspPheHisArgAlaGlyLeuAlaGlu 334
   |||
1081 TTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1120
   |||
334 IleIleAspProIlePheGluPheProArgAlaMetArgArgGluGlyLeu 350
   |||
1141 TCAATTAATTCATTTTCAGTTTTCGTCGGCTATGGGCGGCTGGCTGG 1180
   |||
351 AspAspAlaGluThrAlaLeuLeuLeuAlaIleGluSerAlaAsp 367

```

```

1181 GACGATGTAGACTATGCTTCTGCTATGCTGCTATGCTGCTGCTGCTG 1240
   |||
367 PARGProAsnValGluGluProSerArgValGluAlaLeuGluGluPro 384
   |||
1241 TCGGCTTAATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1280
   |||
384 yValGluAlaLeuLeuSerTyThrArgIleGlySerGlyProIleAspGlu 400
   |||
1281 AGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1340
   |||
401 LeuArgPheProArgMetLeuMetGlySerValSerLeuArgGluLeu 417
   |||
1341 CCGCGCTTCCAGAGATGTCATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
   |||
417 rSerValHisSerGluGluValPheAlaLeuArgLeuHisPylGlySer 434
   |||
1381 CCGCGCTTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
   |||
434 GupProLeuLeuSerGluIleIlePheAspValHisGlu 446
   |||
1441 TGGGCTTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1468
   |||
seq_name: /cqn2_e/prodata/2/123/pe105_02M8.seq:pe1-0595-16311-7
seq_documentation_block:
: sequence 7, Application PC/TUS9516311
: GENERAL INFORMATION:
: APPLICANT: Moore, David
: APPLICANT: Seol, Wonil
: APPLICANT: Choi, Hwang-Sik
: TITLE OF INVENTION: RETINOL X RECEPTOR-INTERACTING
: NUMBER OF SEQUENCES: 17
: CORRESPONDENT ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street, Suite 3100
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2864
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: pe1/0595/16311
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/372,652
: FILING DATE: 13-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Clark, Paul T.
: REGISTRATION NUMBER: 40,162
: REFERENCE/DOCID NUMBER: 00786/246001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1860 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: PCT-0595-16311-7
alignment_scores:
: quality: 2412.00
: ratio: 5.184
: length: 446
: gaps: 0

```


Percent Similarity: 100.000 Percent Identity: 100.000

alignment: 100%

US-09-365-576-3 x PCT-095-16311-7

Align Seq 1/1 for PCT-095-16311-7 from 1 to 1860

```

1 MetSerProThrSerLeuAspThrProValProGlyAsnGlySer 17
141 ATGCTCTGCGGCAACCTGCTGTGACACTGCGTGGTGTGAAAGCTTC 180
17 ProGluProSerThrSerAlaHisSerProThr 1LeuGlyHisGluGly 44
181 TCTGACAGGAGAGACCTGCGGCAACCTGCGGCAACCTGCGGCAAC 240
14 IncludHisProProGluGlySerGluGlySerSerSerAlaPro 50
241 AGGAGACCTGACCTGCTGAGGCTGACAGGCTGACAGGCTGACAGG 280
51 ValVal11111111111111111111111111111111111111 47
281 GTGGTCAACCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 440
67 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 84
431 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
84 ValAlaSerThrHisThrAspValLeuSerCysGluGlyGlySer 100
481 AGGCTTGGGCTTGCACCTACAGAGAGAGAGAGAGAGAGAGAGAGAG 440
441 TTTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
117 GGTGGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144
481 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
144 enuYsArGluAspGlySerGlySerGlySerGlySerGlySerVal 150
541 TGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 580
151 LeuSerGluGluGluGluGluGluGluGluGluGluGluGluGlu 167
581 CTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 640
167 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 184
641 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 680
184 AlaAlaSerProGlyThrSerGluAlaSerSerGluGlySerGlyGly 200
681 GGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 740
201 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 217
741 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
217 TATATATATATATATATATATATATATATATATATATATATATAT 244
781 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
244 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 290
841 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 880
291 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 337
881 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 940
267 GATATATATATATATATATATATATATATATATATATATATATAT 294

```

```

941 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 980
294 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
981 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1040
401 ThrAlaAspThrGlyAsnHisGluGluGluGluGluGluGluGluGlu 317
1041 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
417 ProGluProSerThrSerAlaHisSerProThr 1LeuGlyHisGlu 344
1081 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
444 IncludHisProProGluGlySerGluGlySerSerSerAlaPro 350
1141 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1180
451 AspAspAlaGluGlyAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 367
1181 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1240
467 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1280
1241 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1300
484 YrValGluAlaLeuLeuSerThrThrAlaLeuLeuLeuLeuLeuLeu 400
1281 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1340
417 FserValHisSerGluGluGluGluGluGluGluGluGluGluGluGlu 444
1341 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1400
481 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1460
444 enuProGluProSerGluGluGluGluGluGluGluGluGluGluGlu 446
1441 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500

```

seq: documentation_block:
: Sequence 3, Application US/08342411A
: Patent No. 5639616
: GENERAL INFORMATION:
: APPLICANT: LIAO, Shouling
: APPLICANT: SONG, Chibao
: TITLE OF INVENTION: ORIGINATORS NEURAL RECEPTOR
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESS:
: ADDRESS: Arnold, White & Burke
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: TX
: COUNTRY: USA
: ZIP: 77210 4433
: COMPUTER READABLE FORM:
: MEDIA TYPE: 3.5" 1.44
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC DOS/MS-DOS
: SOFTWARE: Patent in Release #17, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/342 411A
: FILING DATE: 18-Nov 1994
: CLASSIFICATION: 435
: ALTERNATE/OTHER INFORMATION:
: NAME: KITCHILL, BARBARA S.
: REGISTRATION NUMBER: 33,928
: REFERENCE/SEQUENCE NUMBER: Archived

CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER RELEASABLE FORM:
MEDIA TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS 6.22
SOFTWARE: Patent in Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
ACCT DATE: 9-99-99
FILING DATE: Concurrently herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/152,003
FILING DATE: 10-Nov-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HARRIS, S. KITCHEN
REGISTRATION NUMBER: 48,928
REFERENCE/AGENT NUMBER: APT-0540--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-4000
TELEFAX: (713) 787-2679
TELEX: 79-6924
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1959 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MULTIPLE TYPE: DNA (genomic)
POT-US94-12883-4

alignment_scores:

Quality: 2247.50 Length: 416
Ratio: 5.085 Gaps: 1
Percent Similarity: 95.103 Percent Identity: 97.085

alignment_block:

US-09-365-576-3 x POT-US94-12883-4

Align seq 1/1 to: POT-US94-12883-4 from: 1 to: 1959

```
1 MetSerSerProThrSerSerLeuAspHisProValProValProValAspGlySer 17
  |||
256 AIGCTTCGGCTTACAAATTCCTCGCACTCGCTGGCTGGCAATGCTT 405
  |||
17 ProGluProSerThrSerAlaThrSerProThrTleGlySerGluGluGlu 44
  |||
406 TCCGCAATGATGACTGCTTCACTTCAATCAATATTAAGTAAAGTAAATGAT 355
  |||
34 InGluTrAspProProGlySerGluGlySerSerSerAlaGlyGlu 50
  |||
356 AGAAGACTGATTCACCTTCAAGCTCTCAAGAGAGTCCAAATCTGACCAAT 405
  |||
51 ValValIleLeuGluProGluAspGluProGluGluArgGlySer 67
  |||
406 GTG.....GAGCTCAGAGATATGAACTGAGCCCAAGATGCAATATAT 446
  |||
67 ProValProLeuGlySerGluGluGluGluGluGluGluGluGluGluGlu 84
  |||
447 TCCGCGCGGCAAAATGCTGGGCTATGAGCTGTGCGCGCTGTGTGTGGAGGA 496
  |||
84 YSAAlaSerGlyPheHisIleValAsuValIleuSerGlySerGluGlySerGly 100
  |||
497 AGGCTTCGGCTTCCACTATCAATGCTGCTGATGTGAAAGCTGCAAAAG 546
  |||
101 PhePheValMetSerValValIleGlyGlyValGlyValGlyValGlyValGly 117
  |||
547 TCTTTCGGGCTAAGCGTGGTTCATGATGGGCGGCGGCTTAAGCTCTGCTG 596
```

```
217 GATGAGGCTGTTTCTGAGTAAATGATGATGATGATGATGATGATGATGATGAT 144
  |||
597 GAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 646
  |||
124 GAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 150
  |||
617 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 696
  |||
151 LeuSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 167
  |||
627 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 746
  |||
167 nGluProProProProSerGluProAlaAlaSerSerSerGlyArgPro 184
  |||
747 GCAAGCAACGCGCGGAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 796
  |||
184 LalaSerProGlyThrSerGluAlaSerSerSerSerSerSerSerSerSerSer 200
  |||
797 GAGCTGCGCTGAGACCTGCGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 846
  |||
201 GAGCTGCGCTGAGACCTGCGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 217
  |||
847 GAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 896
  |||
217 LAlaAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 234
  |||
897 TATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 946
  |||
234 GTHProProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 250
  |||
947 TCAAGCTGCGCTGAGACCTGCGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 996
  |||
251 GluAlaPheAlaHisPheThrGluGluGluGluGluGluGluGluGluGluGlu 267
  |||
997 GAAAGCTTTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1046
  |||
267 GVALAsPheAlaAlaGluGluGluGluGluGluGluGluGluGluGluGluGlu 284
  |||
1047 GTHProProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1096
  |||
1097 ACCAGATGCGCTGCTGAGGATGATGATGATGATGATGATGATGATGATGATGAT 1146
  |||
401 ThrAlaArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 417
  |||
1147 ACAGCCAGAGCGTTCAACCCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 1196
  |||
417 PheThrTrpSerLysAspAspPheHisAlaGluGluGluGluGluGluGluGlu 434
  |||
1197 GTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1246
  |||
334 IleLeuAsnProIlePheGluPheSerArgAlaAlaGArgArgLeuGluGlu 350
  |||
1247 TCAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1296
  |||
351 AspAspAlaIleTrpAlaLeuGluGluGluGluGluGluGluGluGluGluGlu 367
  |||
1297 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1346
  |||
367 PArgGluAsnValGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 384
  |||
1347 GTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1396
  |||
384 YTVAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 400
  |||
1397 ATGAGAGGCGCTTCCTGCTCAAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAG 1446
  |||
401 LeuValPheProValMetLeuMetLysLeuValSerLeuArgThrLeuSer 417
  |||
1447 GTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1496
```

[illegible]

```

alignment: block:
US-09-365-576-3 x US-08-649-619b-1 ..
Align seq 1/1 101: US-08-649-619b-1 from: 1 to: 1979

1 Met Ser Ser Pro... Thr Ser Ser Leu Asp Thr Pro Val Pro Val Asn Gl
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
206 ATGCTGCTTCTACCAAGAGAGTTCCTGATACGGCTCTGCTGCTGAGAAATGG 255

16 yserProGlnProSerThrSerAlaThrSerProThrIleGlyGlnGlnG 43
1 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
256 CCGCTTCAGATACAGAGAGCTTCTTTCACCTACCTACAGAGAGAGAGAG 305

33 ygdhda.....ThsAspProPro...ProGlySerGly 44
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
306 GCGGAGAGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355

44 GlySerSerSerAlaIleVal.....ValIleLeuGlnAspGlnAs 58
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
356 GAGGCGAGCTACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405

58 pGluProGluArgLysArgLysLysCylGluAlaIleGlySerMetLeuGlyH 75
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
406 GGAATTAAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 455

75 IAsGluLeuGlyArgValLysGlnLysAlaSerGluIlePheHisTyrAsn 91
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
456 AGCAAGCTTTGCCGAGCTGCTGGAGAGAGAGAGAGAGAGAGAGAGAG 505

92 ValLeuSerGlySerGlnIleGlySerLysIlePheProArgAspValValHis 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
506 GAGCGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 555

108 SGILysIleArgLysArgLysArgLysSerGlyThrGlySerHisMetAla 125
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
556 TGGTAAAGGCAAGGCAAGGCTATAGCTGCTGAGAGAGAGAGAGAGAG 605

125 SPALysPheMetArgLysGlySerLeuGlySerArgLysArgLysGlySer 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
606 AGCGCTTATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 655

142 GluAlaGlyMetArgGlnGlnLysValLeuSerGluGlnIleLeuArgLys 158
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
656 GAGCTGAGAGATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 705

158 ArgSerArgIleGlnLysGlnGlnGlnHis.....GlnP 169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
706 GAAAGAGATTCGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 755

169 ProProProSerGlnProAlaAlaLysSerSerSerGluValArgProAlaIle 185
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
756 CAGCTGCTGGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805

186 SerIleProGlyThrSerGlnAlaLysSerGlnGlnLysGlnGlnGlnGln 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
806 TCGCTTGGTGAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 855

202 yIleIleLeuGlyAlaAlaGlnIleGluGlnMetIleGlnGlnGlnValAla 219
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
856 TGTCAGACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 905

219 LAGLLeuGlnLysGlySerLysArgAspGluThrSerArgLysIleProGlySerValThr 235
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
906 CCGAGCTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955

236 ProThrProLeuGlyAlaAspProGlnSerArgAspAlaArgGlnGlnArg 252
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
956 CCGTGGCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005

252 gPheValHisPheThrGluLeuAlaIleIleSerValGlnGlyValAla 269
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1006 CTTTGCTACATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1055

```

[illegible]

```

1  ATTORNEY/AGENT INFORMATION:
2
3  NAME: Dorian, Catherine A.
4  REGISTRATION NUMBER: 36,502
5
6  REFERENCE/WORKET NUMBER: 19416
7
8  ILLUMINATION INFORMATION:
9
10 TELEPHONE: (908) 594-4283
11
12 TELEFAX: (908) 594-4720
13
14 INFORMATION FOR SEQ ID NO: 1:
15
16 SEQUENCE CHARACTERISTICS:
17
18 LENGTH: 2050 base pairs
19
20 TYPE: nucleic acid
21
22 STRANDINESS: single
23
24 FUNCTION: linear
25
26 MOLECULE TYPE: DNA (genomic)
27
28 HYPOTHEetical: NO
29
30 ANTI-SENSE: NO
31
32 US-08-340-518-1

```

aliquant scores =

Quality:	2033.50	Length:	461
Ratio:	4.751	Caps:	5
Percent Similarity:	92.812	Percent Identity:	87.202

```
allignment_block:
```

Align Seq 1/1 to: US-08-440-518-1 from: 1 to: 2030

I MctSes Ser Pro... Thr SerGluAsp ThrProValProCylAsnI 16
| | | | | | | | | | | | | | | | | | |
245 AIGTGTCTTCACAGCAACTCCCTAATACGGATTGCCTGGAAATCG 290

16 ySerPro

295 00000001

3.3.1 YGInn

445 GTCCCGGA

44 GLYCERIN

145 GAGGCCA

58 PG1nPro

445 CCGAACCTTA

15

4.5. CONTACT

References

DATE _____

100-1071

1

A gel electrophoresis image showing a single band in the lane labeled '1'. The band is located at approximately the 100 bp position. The lane is labeled '1' at the bottom.

[illegible]

100

Journal Pre-proof

[illegible]

Quality: 2033.50 Length: 461
 Ratio: 4.751 Gaps: 5
 Percent Similarity: 92.842 Percent Identity: 87.202

alignment_block:

US-09-365-576-3 x US-08-646-248-1 ..

Align seq 1/1 to: US-08-646-248-1 from: 1 to: 2030

```

1 MetSerPro...HisSerLeuAspThrValProGlyAsnG1 16
   |||||
245 MGWVETCTCTAVVAGVAGTTCTGTATACCGCTGGTGTGAATGG 294
   |||||
16 yserProGluProSerThrSerAlaThrSerProThrLysGluGlu 33
   |||||
295 GCTCTTCAAGCTGGAGGCTCTTCTTTCACCCACAGTGAAGAGAGAG 344
   |||||
34 LysThrLeu...ThrAspPro...ProGlySerGlu 43
   |||||
445 GTCGAGAGCGCTGGCTGGAGGCTGGAGCTGATGCTGGAGGACTGAT 394
   |||||
44 GlySerSerSerAlaThrLeuVal...ValThrGluThrGluGlu 58
   |||||
495 GAGGCAATCTCAAGCTTCTCAATGGATCTCAATCTCAATCTCAAT 444
   |||||
58 ProGluProGluGlyAsnGlyLysGlyProAlaProLysMetLeuGly 75
   |||||
445 GCAATCTCAAGGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 494
   |||||
75 LeuThrLeuGlyAsnValLysLysLysLysLysLysLysLysLys 91
   |||||
495 AGCAATCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 544
   |||||
92 ValLeuSerSerGlyGluGlyLysLysLysLysLysLysLysLys 108
   |||||
545 GTCCTCAATCTGGAGGCTGGCTCAAGGCTTCTTCTGGAGCTGGCTG 594
   |||||
108 SerLysLysLysLysLysLysLysLysLysLysLysLysLysLys 135
   |||||
595 TGTGTATATGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 644
   |||||
135 GATATATATATATATATATATATATATATATATATATATATAT 141
   |||||
645 AGCTTTGATATATATATATATATATATATATATATATATATAT 694
   |||||
142 GluAlaGlyMetArgGluThrLysValLeuSerGluGluGluGlu 158
   |||||
695 GAGGCAATCTCAAGGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 744
   |||||
158 SLysAlaGluLeuLysGluGluGluGluGluGluGluGluGlu 169
   |||||
745 GAAVAAATCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 794
   |||||
169 ProProProProSerGluProAlaLysSerSerSerGlyProAla 185
   |||||
795 GAACTTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGG 844
   |||||
186 SerProLysLysSerSerAlaSerSerLysLysLysLysLysLys 202
   |||||
845 TGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 894
   |||||
202 YLLeuLeuLeuThrAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 219
   |||||
895 TGTCTCAATCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 944
   |||||
219 LysLysLysLysLysLysLysLysLysLysLysLysLysLys 235
   |||||
945 TGTCTCAATCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 994
   |||||
240 ProLysProLeuLysLysLysLysLysLysLysLysLysLys 252
   |||||
995 GCTCTTCAAGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGGAG 1044

```

```

252 GPhcAlaHisPheThrGluLeuAlaLeuLeuSerValGluGluLeuValA 269
   |||||
1045 CTCTGGCAGCTTTCAGGAGGCTGGAGGCTGGAGGCTGGAGGCTGGAG 1094
   |||||
269 sPhcAlaLysGluValProGlyPheLeuGluGluGluGluGluGluGlu 285
   |||||
1095 ACTTGGCTTAAGCAGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 1144
   |||||
286 TLeuAlaLeuLysAlaSerThrLeuGluGluGluGluGluGluGlu 302
   |||||
1145 ATCCGCTCTCTGGAGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 1194
   |||||
302 GARGATLysAlaHisSerThrGlyLysLysLysLysLysLysLys 319
   |||||
1195 CAGGCGCTACAGCAGCAGGATCAGGCTGATCAATCTCTCAATGATTTCA 1244
   |||||
319 hTyrSerLysAspAspPheHisArgAlaGluGluGluGluGluGlu 335
   |||||
1245 GCTACAGCAAGGAGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 1294
   |||||
336 AsnProLeuPheGluPheSerArgAlaMetArgArgGluLysLeuAsp 352
   |||||
1295 AACCCAGCTTGGAGCTTCTGGAGGCTGGAGGCTGGAGGCTGGAGGCT 1344
   |||||
352 PAlaGluThrAlaLeuLeuLeuAlaLeuAlaLeuAlaLeuSerAlaAsp 369
   |||||
1345 CCGTCAAGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 1394
   |||||
369 TAsnValGluGluProSerArgValGluGluAlaGluGluGluGluGlu 405
   |||||
1395 CCAAGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGGAG 1444
   |||||
386 GluAlaLeuLeuSerTyrThrArgLysLysArgProGluAspLeuAla 402
   |||||
1445 GAGGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1494
   |||||
402 pPheProArgMetLeuMetLysLeuValSerLysArgThrLysSerGly 419
   |||||
1495 GTTCCGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 1544
   |||||
419 gHisSerGluGluValPheAlaLeuAlaGluGluGluGluGluGlu 435
   |||||
1545 TGCAGCTGGAGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 1594
   |||||
436 ProLeuLeuSerGluLeuLysPheValHisGlu 446
   |||||
1595 CCTCTGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGG 1627
   |||||
seq_name: y09h2.v/prodatay/z/fin/pe105_006b.seq.1.ct 0595-15924-1
seq_documentation_block:
: Sequence 1, Affiliation: pe1059513924
: GENERAL INFORMATION:
: APPLICANT: Friedman, Ethan
: APPLICANT: Holloway, M. Katharine
: APPLICANT: Rodan, Glendon
: APPLICANT: Rutledge, Su Jane
: APPLICANT: Schmidt, Arthur
: APPLICANT: Vogel, Robert
: TITLE OF INVENTION: METHOD FOR FINDING RECEPTOR POTENTIALS
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Morck & Co., Inc.
: STREET: 126 East Lincoln Avenue
: CITY: Rahway
: STATE: New Jersey
: COUNTRY: US
: ZIP: 07065-0907
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```



```

1 CURRENT APPLICATION DATA
2 APPLICATION NUMBER: 1070957/1992
3 FILING DATE:
4 CLASSIFICATION:
5 ATTORNEY/AGENT INFORMATION
6 NAME: Quadagno, Carlo S.
7 REGISTRATION NUMBER: 35,300
8 REFERENCE/PRECEDENCE: 1987 TC
9 TELECOMMUNICATION INFORMATION
10 TELEPHONE: (908) 594-3809
11 TELEFAX: (908) 594-4720
12 INFORMATION FOR SEQ ID NO: 1:
13 SEQUENCE CHARACTERISTICS
14 LENGTH: 2030 base pairs
15 TYPE: nucleic acid
16 STRANDEDNESS: single
17 TOPOLOGY: linear
18 MOLECULE TYPE: DNA (genomic)
19 HYDROLYTICAL: N
20 ANTI-SENSE: N
21 1070957-1/924-1

```

```

Sequence 1: Application: PCT/US95/13931
GENERAL INFORMATION:
  APPLICANT: Frolman, Eitan
  APPLICANT: Hollaway, M. Katharine
  APPLICANT: Rodan, Gideon
  APPLICANT: Schmidt, Aron
  APPLICANT: Vogel, Robert
  TITLE OF INVENTION: USE OF RECEPTOR POTENTIALS
  NUMBER OF SEQUENCES: 5
  CORRESPONDENCE ADDRESS:
  ADDRESS: Motek & Co., Inc.
  STREET: 126 East Lincoln Avenue
  CITY: Rahway
  STATE: New Jersey
  COUNTRY: US
  ZIP: 07065-0907
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patent in Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: PCT/US95/13931
  FILING DATE:
  CLASSIFICATION:
  ATORNEY/AGENT INFORMATION:
  NAME: Qualitro, Carol S.
  REGISTRATION NUMBER: 45,440
  REFERENCE/BOOK NUMBER: 1941b PCT
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (908) 594-3809
  TELEFAX: (908) 594-4720
  INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 2040 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  HYDROLYTIC: NO
  ANTI SENSE: NO
  PCT US95-13931-1

alignment_scores:
  quality: 2033.50          Length: 461
  Ratio: 4.751             Gaps: 5
  Percent Similarity: 92.842 Percent Identity: 87.202

alignment_block:
  US-09-365-576-3 x PCT-US95-13931-1 ...

Align seq 1/1 to: PCT-US95-13931-1 from: 1 to: 2030
1 MotSerSerPro...ThSerSerLeuAspThrProValProGlyAsnG1 16
   |||||
245 AAGGCTCTGCTACGATGATGCTGATGATGATGATGATGATGATGAT 294
   |||||
16 yserProGlyProGlySerValThrSerProThrLeuGlySerGlu 33
   |||||
295 CCGGCTGACGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
   |||||
43 TgcTact...ThAspPro...ProGlySerLeu 43
   |||||
45 GTCGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 394
   |||||
44 GlySerSerAlaTyrLeVal...ValLeuGluGluProGluAs 58
   |||||
495 GAGGCAACCTGACCTTGTAAACACACCTTGTGACCTTGTGACCTG 444
   |||||
56 pGluProGluAlaGlyAlaGlySerLeuGluAlaProGlyMetLeuGly 75

```

```

|||||
445 GAGACCAAGGCTAAGGCAGAAAGAGGGGCAACCTGCAAGATGCTGGCG 494
   |||||
75 TSGTuleGysAr4ValGysGlyAspGlyAlaSerGlyProHisIeIAsp 91
   |||||
495 ATGAGCTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 544
   |||||
92 ValLeuSerGlySerGluGlySerGlySerGlySerGlySerGlySer 108
   |||||
545 GTCGCTACGCTGCAAGGCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCT 594
   |||||
108 SGLGGLAAGLAAATGATGATGATGATGATGATGATGATGATGATGAT 125
   |||||
545 TGGTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 644
   |||||
125 SPALaHeKtATATATATATATATATATATATATATATATATATAT 141
   |||||
645 ACGCTTCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 694
   |||||
142 GUAAGAGLymeATAGLuglGlyValLeuSerGluGluGluLeuAla 158
   |||||
695 GAGGCAAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 744
   |||||
158 slyAsqLleGluGluGluGluGluGluGluGluGluGluGluGluGlu 169
   |||||
745 GAGGAGATTCGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 794
   |||||
169 yserProGlySerGlyProAlaIleSerSerSerGlyValProAla 185
   |||||
795 CAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844
   |||||
186 SerProGlyThrSerGluAlaSerSerGluGluGluGluGluGluGlu 202
   |||||
845 TCCCGCTGGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 894
   |||||
202 ylleGluGluThrAlaAlaGluGluGluGluGluGluGluGluGlu 219
   |||||
895 TGTCAGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 944
   |||||
219 IagGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 235
   |||||
945 CTCAGCTGAGTGCAGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 994
   |||||
236 ProTrpProLeuGlyAlaAspProIleSerValAlaIleArgGluAla 252
   |||||
995 CCGTGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1044
   |||||
252 GPhenAlaHisProThrGluGluAlaIleIleSerValGluGluValAla 269
   |||||
1045 CTTGGCTGAGCTTGCAGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1094
   |||||
269 SPPhenAlaGlySerValProGlyProGluGluGluGluGluGluGlu 285
   |||||
1095 ACTTGCTAAGCAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1144
   |||||
286 TleAlaLeuLeuGlyAlaSerThrTleGluIleLeuGluGluGluThrAl 302
   |||||
1145 ATGGCCCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1194
   |||||
302 AAGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 319
   |||||
1195 CAGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1244
   |||||
319 hTrpSerLeuGlyAspProHisAlaGluGluGluGluGluGluGlu 335
   |||||
1245 CTACAGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1294
   |||||
336 AspProIleLeuGluGluGluSerAlaAlaLeuAlaGluGluGluGlu 352
   |||||
1295 AACCCGATCTTGTGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1344
   |||||
352 pAlaGluTyrAlaLeuLeuIleAlaIleAsnIleGlySerAlaAspAla 369
   |||||

```

```
seq_documentation_block,
; Sequence 1, Application US/08342411A
; ...
; ...
```

alignment_scores:

[illegible][illegible]

IMMEDIATE SOURCE:
CLONE: XP2 (XP2.10C)
FEATURE:
NAME/NET: CDS
LOCATION: 148..1470
US-08-444-358-7

alignment_scores:
Quality: 1495.00 Length: 450
Ratio: 4.443 Gaps: 7
Percent Similarity: 80.567 Percent Identity: 63.111

alignment_block:

US-09-365-576-3 x US-08-444-358-7

Align seq 1/1 10: US-08 334-458-7 from 1 10: 1659

```
4 ProhrSerSerTouASPTrProValProGly...ASuLysSerProG 19
111 |-----|-----|-----|-----|-----|-----|
187 CCTAACTTGGGAGTGGAGCTGTGGAGCCAGAGAGAGAGAGAGAGAG 246
19 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 36
111 |-----|-----|-----|-----|-----|-----|
237 CCTAACTTGGGAGTGGAGCTGTGGAGCCAGAGAGAGAGAGAGAGAG 282
46 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 50
111 |-----|-----|-----|-----|-----|-----|
283 |-----|-----|-----|-----|-----|-----|
51 ValValLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 64
111 |-----|-----|-----|-----|-----|-----|
425 CTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 474
63 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 80
111 |-----|-----|-----|-----|-----|-----|
475 CCTAACTTGGGAGTGGAGCTGTGGAGCCAGAGAGAGAGAGAGAGAG 424
80 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 96
111 |-----|-----|-----|-----|-----|-----|
425 CTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 474
97 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 111
111 |-----|-----|-----|-----|-----|-----|
475 CCTAACTTGGGAGTGGAGCTGTGGAGCCAGAGAGAGAGAGAGAGAG 518
113 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 130
111 |-----|-----|-----|-----|-----|-----|
519 CTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 568
130 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 146
111 |-----|-----|-----|-----|-----|-----|
569 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 618
147 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 163
111 |-----|-----|-----|-----|-----|-----|
619 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 668
163 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 180
111 |-----|-----|-----|-----|-----|-----|
669 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 718
180 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 196
111 |-----|-----|-----|-----|-----|-----|
719 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 738
197 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 213
111 |-----|-----|-----|-----|-----|-----|
749 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 767
213 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 230
111 |-----|-----|-----|-----|-----|-----|
768 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 817
```

```
230 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 246
111 |-----|-----|-----|-----|-----|-----|
818 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 867
247 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 263
111 |-----|-----|-----|-----|-----|-----|
868 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 917
263 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 280
111 |-----|-----|-----|-----|-----|-----|
918 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967
280 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 296
111 |-----|-----|-----|-----|-----|-----|
968 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017
297 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 313
111 |-----|-----|-----|-----|-----|-----|
1018 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1067
413 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 330
111 |-----|-----|-----|-----|-----|-----|
1068 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
430 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 346
111 |-----|-----|-----|-----|-----|-----|
1118 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1167
447 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 363
111 |-----|-----|-----|-----|-----|-----|
1168 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1217
463 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 380
111 |-----|-----|-----|-----|-----|-----|
1218 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267
480 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 396
111 |-----|-----|-----|-----|-----|-----|
1268 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
497 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 413
111 |-----|-----|-----|-----|-----|-----|
1318 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1367
413 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 430
111 |-----|-----|-----|-----|-----|-----|
1368 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
430 hProhrSerSerTouASPTrProValProGly...ASuLysSerProG 446
111 |-----|-----|-----|-----|-----|-----|
1418 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1467
```

seq_name: us-09-365-576-3, seq_id: 1, seq_len: 1659, seq_start: 1, seq_end: 1659

seq_documentation_block:

Sequence 1: Application 95/09463694

General Information:

Patent No. 5696233

Applicant: EVANS Ph.D., RONALD M.

Applicant: MANGELSDORF Ph.D., DAVID J.

Applicant: ONE MS., ESTHER A. S.

Applicant: ONE Ph.D., ANTHONY E.

Applicant: BOEKMEYER Ph.D., OWEN K.

Applicant: GIGHERIE Ph.D., VICTOR M.

Applicant: YAO M., ISO-PANG MNM

Title of Invention: REVEL RESEARCH

Number of Sequences: 14

Correspondence Address:

Address: Preilly, Schroeder, Brueneman & Clark

Street: 444 So. Flower St., Suite 2000

City: Los Angeles

State: CA

COUNTRY: US
ZAP: 90071-2921
MODIFIER: READABLE FORM
MEDIUM TYPE: Floppy disk
COMPUTER: IBM pc compatible
OPERATING SYSTEM: pc-rtos/386-rtos
SOFTWARE: Patient In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,694
FILING DATE: 05-JUN 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/761,063
FILING DATE: 17-SEP 1991
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 41192
REFERENCE/PATENT NUMBER: P41 8939
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 545-9001
TELEFAX: (619) 545-8949
INFORMATION POP-UP ID NO: 73
SEQUENCE CHARACTERISTICS:
LENGTH: 1659 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MULTIPLE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: XR2 (XR2-Seq)
FEATURE:
NAME/KEY: cDS
LOCATION: 148..1476
US-08-464-694-7

alignment scores:
Quality: 1395.00 Length: 450
Ratio: 3.84 Gaps: 7
Percent Similarity: 80.667 Percent Identity: 63.111

alignment block:
US-09-365-576-3 x US-08-464-694-7
Align seq 1/1 for US-08-464-694-7 from 1 to 1659

4 proteinSerLeuAspThrProValProGly...AsnGlySerProG1 19
117 TTTGACTTCTGATGGAATCGTGCAGATTCAGAGATGAAAG 246
187 TTTGACTTCTGATGGAATCGTGCAGATTCAGAGATGAAAG 246
19 outSerThrSerAlaAlaSerProThrTleGlySerGlnGlyGlnG1 36
237 CTAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 282
46 InAspProThrProGlySerGlnGlySerSerSer...AlaGlyTle 50
283ATGATTCATCTCTGAGGCTTCAAGGCTTCAAGGCTTCAAG 324
51 ValValTleLeuGlnProGlnAspGlnPro.....GlnAcGly 63
425 TTTGACTTCTGATGGAATCGTGCAGATTCAGAGATTCAGAGATTCAG 374
63 sATGATTCATCTCTGAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 374
375 GTCGAAAAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 424
80 ATGATTCATCTCTGAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 474
425 TTTGATTCATCTCTGAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 474
97 GTCATTCATCTCTGAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 514
475 GTCATTCATCTCTGAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 514

113 gTATATGysATGlySerGlyThrGlyGlnMetAspAlaThrMetArg 140
117 TTTGACTTCTGATGGAATCGTGCAGATTCAGAGATTCAGAGATTCAG 187
519 CTACATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 568
130 TATGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 146
559 GTCATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 618
147 GlnGlnGysValLeuSerGlnGlnGlnLeuGlyGlySerArgTleGln 163
619 GATGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 668
163 SerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 180
669 GATGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 718
180 GTATATGTCATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTC 196
719 CA...GTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 748
197 SerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 214
749GATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCAG 767
213 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 240
768 GATGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 817
230 SerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 246
818 ACCGCTTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 867
247 AspAlaThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 263
868 GATGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 917
263 TATGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280
948 TGTGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 967
290 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 296
968 TATGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1017
297 MetLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 313
1018 ATGCTTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 1067
313 rPhGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 440
1068 CTTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTC 1117
330 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 446
1118 TGTGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 1167
347 ArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 363
1168 GATGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 1217
363 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 400
1218 CTTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTC 1267
380 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 446
1268 TGTGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 1317
397 ProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 413
1318 GTCATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAGATTCGTCAG 1367


```
413  UATGHIILeulSerSerValHisSerThiGlnValIleAlaIleuArgLeuG 440
      |||
1468  CTGGATCCCTGGACAGCGCTATCTAGAGGCAAGTGGTGGTCTGGTCTGG 1417
      |||
430  InAspLysLysLeuProIleuLeuSerGluIleTTPAspValHisGlu 446
      |||
1418  AGGATCAAAAAGCTCCGACGCGCTGCTCTCTGACATCTGGATCTGCACCGAA 1467
```



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run On: April 29, 2001, 14:17:27 : Search Time 2399.16 Seconds

(without alignments)
4792.079 Million cell updates/sec

Title: US-09-365-576-7

Percent score: 1860
Sequence: 1 GGCAGAGGACACAGCTGCA.....AAAAAAAAAAAAAAAAAAAA 1860

Scoring table: IDENTITY_ND2
Gapop 10.0 , Gapext 1.0

Searched: 128325 seqs, 3078962356 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```

GenEm1:*
1: qb_bal:*
2: qb_bal:*
3: qb_bal:*
4: qb_bal:*
5: qb_bal:*
6: qb_bal:*
7: qb_bal:*
8: qb_bal:*
9: qb_bal:*
10: qb_bal:*
11: qb_bal:*
12: qb_bal:*
13: qb_bal:*
14: qb_bal:*
15: qb_bal:*
16: qb_bal:*
17: qb_bal:*
18: qb_bal:*
19: qb_bal:*
20: qb_bal:*
21: qb_bal:*
22: qb_bal:*
23: qb_bal:*
24: qb_bal:*
25: qb_bal:*
26: qb_bal:*
27: qb_bal:*
28: qb_bal:*
29: qb_bal:*
30: qb_bal:*
31: qb_bal:*
32: qb_bal:*
33: qb_bal:*
34: qb_bal:*
35: qb_bal:*
36: qb_bal:*
37: qb_bal:*
38: qb_bal:*
39: qb_bal:*
40: qb_bal:*
41: qb_bal:*
42: qb_bal:*
43: qb_bal:*

```

```

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: qb_sts1:*
54: qb_sts2:*
55: qb_sts3:*
56: qb_sy:*
57: qb_un:*
58: qb_vl:*
59: qb_vl2:*
60: qb_vl3:*
61: qb_vl4:*
62: qb_vl5:*
63: qb_vl6:*
64: qb_vl7:*
65: qb_vl8:*
66: qb_vl9:*
67: qb_vl10:*
68: qb_vl11:*
69: qb_vl12:*
70: qb_vl13:*
71: qb_vl14:*
72: qb_vl15:*
73: qb_vl16:*
74: qb_vl17:*
75: qb_vl18:*
76: qb_vl19:*
77: qb_vl20:*
78: qb_vl21:*
79: qb_vl22:*
80: qb_vl23:*
81: qb_vl24:*
82: qb_vl25:*
83: qb_vl26:*
84: qb_vl27:*
85: qb_vl28:*
86: qb_vl29:*
87: qb_vl30:*
88: qb_vl31:*
89: qb_vl32:*
90: qb_vl33:*
91: qb_vl34:*
92: qb_vl35:*
93: qb_vl36:*
94: qb_vl37:*
95: qb_vl38:*

```

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1877.4	98.9	1841	94	MM000419
2	1589.2	85.4	1934	9	A49139 Sequence 1
3	1569	84.4	1877	95	PM020386
4	1542.6	82.9	1959	10	146766 Sequence 1
5	1542.6	82.9	1959	95	EM034575
6	1025	55.1	1898	10	146765
7	1023	55.0	1979	9	AK035516 Sequence
8	1023	55.0	1979	45	E11456
9	1022.2	55.0	2030	10	136447
10	1022.2	55.0	2030	10	179211
11	1019.8	54.8	2010	93	HSD07112

Db	1917	AA	1918	
Qy	1859	AA	1860	
Db	1857	TTTATT	TTAAATAAAAAAT	AAAAATAAAAA
Qy	1799	TTTATT	TTAATAAAAAA	AAAAATAAAAA
Db	1797	AGAGCT	TAGAGCGAGGTGGCT	CGCCACAGTGGAGAGACGAGGCGCTCTTGGCCCT
Qy	1739	ACAGCGT	TAGAGCGAGGTGGCT	CGCCACAGTGGAGAGACGAGGCGCTCTTGGCCCT
Db	1737	CAGTGGCGCTTGGCTGGAGGCT	TAAACGCTGACGCTTACACGCTACACGCTACGCTTACGCTGAG	1736
Qy	1679	CAGTGGCGCTTGGCTGGAGGCT	TAAACGCTGACGCTTACACGCTACACGCTACGCTTACGCTGAG	1738
Db	1637	CAGTGGCGCTTGGCTGGAGGCT	TAAACGCTGACGCTTACACGCTACACGCTACGCTTACGCTGAG	1796
Qy	1619	CCCAACCATTTGATCTCTCTTAAAGATGTTTATAGGCTTAAAGGCTCTGAGGCTTAC		1678
Db	1617	CCCAACCATTTGATCTCTCTTAAAGATGTTTATAGGCTTAAAGGCTCTGAGGCTTAC		1736
Qy	1649	ACGCTTAAATCTCTCTTAAAGATGTTTATAGGCTTAAAGGCTCTGAGGCTTAC		1676

[illegible]

REFERENCE : (bases 1 to 215601)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Ref: Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 215601)
Ref: Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, Ref: Joint
Genome Institute, 2900 Mitchell Drive, Walnut Creek, CA 94598, USA
on Jul 18, 2000. This sequence version replaced 115919423.

Genome Center
Center, Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1755238
Center clone name: R97-23_4786

Summary Statistics

Consensus quality: 20615 bases at least Q40
Consensus quality: 21090 bases at least Q30
Consensus quality: 214651 bases at least Q20
Estimated insert size: 220350; adaptor-tip estimation
Estimated insert size: 214601; sum-of-coverage estimation
Quality coverage: 7.98 in Q20 bases; adaptor-tip estimation
Quality coverage: 8.1 in Q20 bases; sum-of-coverage estimation.
NOTE: This is a 'working draft' sequence. It currently

* consists of 22 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 4864: contig of 4864 bp in length
* 4865 4864: gap of unknown length
* 4965 21161: contig of 16197 bp in length
* 21162 21161: gap of unknown length
* 21462 22873: contig of 812 bp in length
* 22074 22173: gap of unknown length
* 22174 43484: contig of 11311 bp in length
* 4485 43584: gap of unknown length
* 4485 49528: contig of 5044 bp in length
* 49529 39628: gap of unknown length
* 49629 62157: contig of 22523 bp in length
* 62158 62257: gap of unknown length
* 62258 76745: contig of 14488 bp in length
* 76746 76845: gap of unknown length
* 76846 81102: contig of 4257 bp in length
* 81103 81102: gap of unknown length
* 81203 85706: contig of 4504 bp in length
* 85707 85806: gap of unknown length
* 85807 90755: contig of 4949 bp in length
* 90756 90855: gap of unknown length
* 90856 101543: contig of 10688 bp in length
* 101544 101643: gap of unknown length
* 101644 119428: contig of 17685 bp in length
* 119429 119428: gap of unknown length
* 119429 122689: contig of 4261 bp in length
* 122690 122789: gap of unknown length
* 122790 131818: contig of 9329 bp in length
* 131819 131918: gap of unknown length
* 131919 150107: contig of 18189 bp in length
* 150108 150207: gap of unknown length
* 150208 156868: contig of 6661 bp in length
* 156869 156968: gap of unknown length
* 156969 187233: contig of 30755 bp in length
* 187234 187823: gap of unknown length
* 187824 190416: contig of 2593 bp in length
* 190417 195946: gap of unknown length
* 195947 196056: contig of 5420 bp in length
* 196057 196056: gap of unknown length

FEATURES
Source
1. 215601
/organism: "Mus musculus"
/db_xref: "taxon:10090"
/feature: "Ref23_4786"
/feature: "Ref23_4786"
BASE COUNT 52009 a 52880 c 58490 g 59020 t 2101 others
ORIGIN

Query Match 29,136 Score 541.4; Dh 75; Length 215601;
Best Local Similarity 81.1%; Prev. No. 8,790-196;
Matches 731; Conservative 0; Mismatches 1; Indels 169; Gaps 2;

QY 1109 GCTGAGGCTTCAGTGGGAAATTCATGCAATTCGATGCTGGGAGGATGAG 1168
1109 GCTGAGGCTTCAGTGGGAAATTCATGCAATTCGATGCTGGGAGGATGAG 1168
DB 11408 GCTGAGGCTTCAGTGGGAAATTCATGCAATTCGATGCTGGGAGGATGAG 114142
11408 GCTGAGGCTTCAGTGGGAAATTCATGCAATTCGATGCTGGGAGGATGAG 114142
QY 1169 GAGTGGGCTTGGAGATGCAAGATAGCTTGGTATGCTATCAACATCTGAG 11228
1169 GAGTGGGCTTGGAGATGCAAGATAGCTTGGTATGCTATCAACATCTGAG 11228
DB 11414 GAGTGGGCTTGGAGATGCAAGATAGCTTGGTATGCTATCAACATCTGAG 114202
11414 GAGTGGGCTTGGAGATGCAAGATAGCTTGGTATGCTATCAACATCTGAG 114202
QY 1229 GATGGGCTTAAATGCTGAGGATGCAAGATGCTGAGGATGCTGAGGATGCTGAG 1288
1229 GATGGGCTTAAATGCTGAGGATGCAAGATGCTGAGGATGCTGAGGATGCTGAG 1288
DB 11429 GATGGGCTTAAATGCTGAGGATGCAAGATGCTGAGGATGCTGAGGATGCTGAG 114292
11429 GATGGGCTTAAATGCTGAGGATGCAAGATGCTGAGGATGCTGAGGATGCTGAG 114292
QY 1289 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1421
1289 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1421
DB 11426 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114422
11426 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114422
QY 1322 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1421
1322 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1421
DB 11438 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114442
11438 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114442
QY 1322 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1460
1322 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1460
DB 11444 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114502
11444 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114502
QY 1361 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1420
1361 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1420
DB 11450 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114562
11450 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114562
QY 1421 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1480
1421 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1480
DB 11456 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114622
11456 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114622
QY 1481 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1540
1481 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1540
DB 11462 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114482
11462 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114482
QY 1541 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1600
1541 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1600
DB 11468 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114742
11468 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114742
QY 1601 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1660
1601 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1660
DB 11474 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114801
11474 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114801
QY 1661 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1720
1661 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1720
DB 11480 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114861
11480 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114861
QY 1721 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1780
1721 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1780

Id 114862 CATGACCTTGATTGAGAGACTTAGGGCAATGATGCGGCTAAATTTGGGAGATTAAG 114921
QY 1781 GCGCTCTCTTCTGCGGCTTTATTTATTAIAAAAAATAAAAATAAATTTGASTATAGSTC 1840
Id 114922 GCGGCTCTCTCTGCGGCTTTATTTATTAIAAAAAATAAAAATAAATTTGASTATAGSTC 114981
QY 1841 A 1842
Id 114982 A 114982

Search completed: April 29, 2001, 14:16:52
Job Time: 7165 Sec

•
•
•

Query Match	55.18;	Score 1025;	DB 18;	Length 1898;
Best Local Similarity	77.38;	Pred. No. 2.9e+008;		
Matches 1399;	Conservative	0;	Mismatches 345;	Indels 67;
				Gaps 10;

[illegible]

UY	449	GTTCACGCGTGGGGGCGGTAAAGCTGTGTCGGGCAAGGATGGCAGATGATAGT	568
Dh	590	TT	649
UY	509	TTTCATGCGCGCAAGTGTGCAAGCTTGGCGGCTGTGTAAAGTGCAAGCACTTGGCAGCAG	568
Dh	650	TTCTATGCGGCGCAATTCGCAATGTGTGCGGTCTGTCAATGTGTCAATGTGTGTGT	709
UY	569	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	628
Dh	710	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	769
UY	629	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	678
Dh	770	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	829
UY	674	---GGGTTAGTGGGTTCTCTTGGTAACTTGGAGAGAGAGAGAGAGAGAGAG	730
Dh	830	TTTATGCTGGGCTCTCTGTGTGATCTGTGAGCGAGAGAGAGAGAGAGAGAG	889
UY	741	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	790
Dh	890	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	949
UY	791	GTGAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	850
Dh	950	CTGAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1009
UY	851	GTGAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	910
Dh	1010	GTGAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1069
UY	911	ATCATGTGGGTGTGAGAGATTTGAGATTTGTAAATAGTGTGTGTGTGTGTGT	970
Dh	1070	ATCATGTGGGTGTGAGAGATTTGAGATTTGTAAATAGTGTGTGTGTGTGTGT	1129
UY	971	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	1030
Dh	1130	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	1189
UY	1031	ACAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1090
Dh	1190	ACAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1249
UY	1091	AGCAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1150
Dh	1250	AGCAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1309
UY	1151	TTTCATGCGCGCAAGTGTGCAAGCTTGGCGGCTGTGTAAAGTGCAAGCACTTGGCAGCAG	1211
Dh	1310	TTTCATGCGCGCAAGTGTGCAAGCTTGGCGGCTGTGTAAAGTGCAAGCACTTGGCAGCAG	1369
UY	1211	ATTCAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1270
Dh	1370	ATTCAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1429
UY	1271	CAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	1330
Dh	1430	CAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	1489
UY	1331	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	1390
Dh	1490	GAGTAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	1549
UY	1391	TGGAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	1450
Dh	1550	TGGAGTGTGTGCTCTTGATAGTAAATTGGATAAAGAGATTACAGAGAGTAAAG	1609
UY	1451	ATTCAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1501
Dh	1610	ATTCAGTGTGAGCAAAAGCAATTTCTTGAGAGAGAGAGAGAGAGAGAGAGAG	1669

[illegible]

XX	I27616;
XX	06-NOV-1996 (first entry)
PT	
XX	Human foetal lung steroid hormone receptor analogue ECDN cDNA.
LE	
XX	
KM	Human; foetal lung; steroid hormone; receptor; analogue protein;
KM	ECDN protein; cancer; screening; binding molecule; recombinant;
KM	identification; anticancer drug; cancerous tissue; primer;
KM	probe; antibody; immunohistochemical assay; ds.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
PH	206..1591
FT	CDS
ET	/ftaa - a
PN	W09609324-A1.
PD	
PD	28-MAR-1996.
XX	
XX	21-SEP-1995:
XX	95MKO-JP01509.
XX	
XX	21-SEP-1994:
XX	94JP-0226270.
PA	(EISA) EISAI CO LTD.
PA	(GANK-) ZH CAN KENRYUKAI.
PA	(CANC-) CANCER INST.
XX	
P1	Nakamura Y, Saito H;
XX	
DK	WPI; 1996 188403/19.
DR	P-PDB: R96234.
XX	
PT	ECDN protein, a steroid hormone receptor analogue from human foetal
PT	lung - is expressed in cancer cells and is useful for cancer
PT	diagnosis and drug development
XX	
TS	Claim 1; Pages 21-25; 4pp; Japanese.
XX	
CC	The present sequence encodes the human foetal lung derived steroid
CC	hormone receptor analogue protein, ECDN. In various cancer cells a
CC	variant ECDN protein, designated ECDN small mol. (ECINsm) protein,
CC	is expressed. Therefore screening for ECDN and ECINsm protein
CC	binding mols., using recombinant ECDN and ECINsm proteins will be
CC	useful in the identification of candidate anticancer drugs. Gene
CC	expression of ECDN and ECINsm proteins in normal and cancerous
CC	tissues can be studied using primers and probes derived from ECDN
CC	and ECINsm protein cDNA. Antibodies which recognise ECDN and ECINsm
CC	proteins can be used in ECDN and ECINsm protein
CC	immunohistochemical assays.
XX	
SQ	Sequence 1979 HQ: 405 A; 642 C; 602 G; 330 T; 0 other;
XX	
Query Match	54.9%; Score 1021.4; Off 17; Length 1979;
Inst. Local Similarity	77.0%; Prev. No. 26-227;
Matches 1424;	Conservative 0; Mismatches 356; Indels 69; Gaps 12;
Y	46 CGGGCGTCAGTATACGCCTGCTGTGTTGTTGTGTAATGACGAGCGATGGTGTTGGAA 105
D	T T
D	123 GCGAGTGTGAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 182
Y	106 CAGCACGGCTGGCTTTCGTAACCGACACTATG-TCTTGGCGCAAGAATTCTGTGAACATGCC 164
D	T T
Y	183 AGAGGTGTCTCCCTGTGACCCACCATCTCTCTCTACCAACAATTCTCTGTATATATCCC 242
Y	165 TGCTGGGAAAAGTGTCTGTCTGACCGATATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTG 224
D	T T
D	243 TGCTTGAATAATGTGCT 402
Y	225 ACCGCGTACGACATCATGTC-----GTCTTGTGACGGCGGTGAAAGCTGTGCA 206

QY 1786 -CGCTGTCGCGCTTATTAATAAAAACAAATAATTTGAGTACAAAGTAAAA 1844
 DB 1470 -CGCTGTCGCGCTTATTAATAAAAACAAATAATTTGAGTACAAAGTAAAA 2029
 QY 1845 A 1845
 DB 2030 1 2030
 RESULT 10
 Q88760
 ID Q88760 standard: DNA: 1813 bp.
 XX
 AC Q88760:
 XX
 DT 20-JAN-1996 (first entry)
 XX
 DE Human ubiquitous nuclear receptor protein DNA.
 XX
 KW Ubiquitous nuclear receptor; drug design; disease diagnosis;
 KW therapy; se.
 XX
 OS Homo sapiens
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1727
 FT CDS /frame -a
 XX
 FN W09513373 A1.
 XX
 PD 18-MAY-1995
 PE 08-NOV-1994: 6460-US12484.
 XX
 PR 10-NOV-1993: 9305-0152003.
 XX
 PA (ARCH-) ARCH DNV CORP
 XX
 PI Liao S, Song C;
 XX
 DE WPI 1996 141092/26
 DE P-PSDB: R74738.
 XX
 PT New ubiquitous nuclear receptor - used to develop prods. for use in
 PT diagnostic drug design and therapeutic applications
 XX
 PS Claim 6: Page 148; 1sep97; English.
 PS This genomic DNA isolated from human prostate cancer pro-3 cells may
 CC be expressed recombinantly to produce a UNRP which interacts with
 CC the response elements and network of receptors in the thyroid
 CC hormone/thyroid acid receptor subfamily. The products can be used
 CC in detection, diagnosis and in screening assays for substances which
 CC interact with UNRPs for use in diagnosis, drug design and
 CC therapeutic applications
 CC
 SO Sequence 1813 bp; 410 A; 573 G; 529 C; 301 T; 0 other

Query Match 54.4% Score 10127 DB 16 Length 1813
 Best local similarity 77.4% Prod NO. 2626
 Matches 1380; Conservative 0; Mismatches 325; Indels 67; Gaps 10;

QY 146 AGTCTCTGACACTCGCGCTGCTGAGATGGTCTGCTCAAGGACAGTCTGCTGACG 205
 DB 1 agtctctgatacagcccttgccttgaatagcccccctcagcttgagctctctctt 60
 QY 206 TCACCTCTTATACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 247
 DB 61 ttagctgctcttataag 120
 QY 248 CCAAGCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301
 DB 111

DB 121 cag 180
 QY 302 GATGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 361
 DB 181 gag 240
 QY 362 TGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 421
 DB 241 tggag 300
 QY 422 TGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 481
 DB 301 cag 460
 QY 482 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 541
 DB 361 ggtgag 420
 QY 542 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 601
 DB 421 ggtgag 480
 QY 602 AATGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
 DB 481 aatgag 540
 QY 661 -----GAGGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720
 DB 541 ggtgag 600
 QY 704 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 763
 DB 601 ggtgag 660
 QY 764 AATGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 823
 DB 661 atgag 720
 QY 824 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 883
 DB 721 ggtgag 780
 QY 884 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 943
 DB 781 ggtgag 840
 QY 944 AATGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1003
 DB 841 aatgag 900
 QY 1004 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1063
 DB 901 ggtgag 960
 QY 1064 AATGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1123
 DB 961 aatgag 1020
 QY 1124 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1183
 DB 1021 ggtgag 1080
 QY 1184 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1243
 DB 1081 ggtgag 1140
 QY 1244 GGTGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1303
 DB 1141 ggtgag 1200
 QY 1304 AATGAGGCTGAGTGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1363
 DB 1201 aatgag 1260

UY	899	ATGACCTAGCAATGCAATGCGCGCTGAGAGAGATTTGGCATTTGCTGCAACCGGTCGCACGG	958
Db	728	atcgagagctggcattccatccatccagagagatcgaaggaattccgttcaatcaagagagcctgggt	787
UY	959	TCTCTGCAATTTGCGCGCGGAGAGCAATGAGATGCGCTGCTGCAACGCTGCTGCAATTGAGATC	1018
Db	788	ttctcttgcgaattggcgccgagagagagccggaatcccccctccgaagcgatccaatccagatccgagtc	847
UY	1019	ATGCTTCTTCAAAATGAGCAAGCACTTCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1078
Db	848	atgctctgctatag	907
UY	1079	GACTTCAATCAATGAGCAAGCAAGCAATTCACCGGTCGACGGCTTCGACAGCTGGATATCATCAT	1138
Db	908	gaattctgagctacag	967
UY	1139	ATGCAATTTGCAATTTGCGCGCGGAGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1198
Db	968	ctgctatcttgcgaattctcgag	1027
UY	1199	TGCTCTTCAATCAATCAATCAATCTCTTCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1258
Db	1028	ctgctatcttgcgaattctcgag	1087
UY	1259	GTCGACAGCGCTGAGCAAGCGCTTCAGGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1318
Db	1088	gtgag	1147
UY	1319	ATGCAATTTGCAATTTGCGCGCGGAGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1378
Db	1148	ctgctatcttgcgaattctcgag	1207
UY	1379	ATGCAATTTGCAATTTGCGCGCGGAGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1438
Db	1208	gtgag	1267
UY	1439	TGCTCTTCAATCAATCAATCAATCTCTTCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1498
Db	1268	gtgag	1327
UY	1466	ATGCAATTTGCAATTTGCGCGCGGAGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1526
Db	1328	gtgag	1387
UY	1553	TGCTCTTCAATCAATCAATCAATCTCTTCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1600
Db	1468	gtgag	1447
UY	1601	ATGCAATTTGCAATTTGCGCGCGGAGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1658
Db	1448	gtgag	1507
UY	1659	TGCTCTTCAATCAATCAATCAATCTCTTCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1718
Db	1508	gtgag	1567
UY	1718	ATGCAATTTGCAATTTGCGCGCGGAGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1773
Db	1568	gtgag	1627
UY	1774	ATGCAATTTGCAATTTGCGCGCGGAGAGCAATGAGCAATGAGCAATGAGCAATGAGCAATGAGCAAT	1825
Db	1628	gtgag	1680
RESEQUENCE	13		
UY	235343		
XX	Z35043	standard; cDNA; 1528 bp.	
XX	Z35043		
XX	28-FEB-2000	(first entry)	

db 463 ggaagcagcctatgctctgacacagcgttgatgacatgcccacagatgacatgacatgacgag 522
uy 521 AAGTGGCAATCTCTGCGAGCTGCGTGAAGTGCAGAGAGAGCTGCGTATGCGAGAGAGTGGCTG 580
db 523 aatggcagcagatgtgactctgcaaatgacagcagcagcagcagcagcagcagcagcagcagc 582
uy 581 CTCTCTTGAAGTATATTTGGAGAGAGAGATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 640
db 583 ctgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 642
uy 641 CCGCCATCTTGAAG 700
db 643 ggcaccltctgctgctcgcacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 689
uy 701 GAAGCAAG 760
db 690 -----acagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 711
uy 761 CTGATGATCTGAT 820
db 712 ggcataatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 771
uy 821 CAAGCCCAAGTATAG 880
db 772 cgcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 831
uy 881 CAAGCTTTGCTAT 940
db 832 cagc 891
uy 941 GTCAGTATCTGAT 1000
db 892 gccaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 951
uy 1001 GCGTGCATCTGAT 1060
db 952 acctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1011
uy 1061 TGATATAGTAT 1120
db 1012 aacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1071
uy 1121 CAGGTGCAATCTGAT 1180
db 1072 caggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1131
uy 1181 GACGATATAG 1240
db 1132 aatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1191
uy 1241 GTGAG 1300
db 1192 gtagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1251
uy 1301 TACAG 1360
db 1252 taagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1311
uy 1361 GTGAG 1420
db 1312 gtagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1371
uy 1421 GACAG 1469
db 1372 gacaaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1420

3
4
5
6

[illegible]

ADDRESSEE: Motec S Co., Inc.
STREET: 126 East Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patient's Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
AGE: FIVE TEN THREE-08/02/94 11:23H
FILING DATE: 14-MAY-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/340,284
FILING DATE: 27-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Dolan, Catherine A.
REGISTRATION NUMBER: 36,502
REFERENCE/DOCKET NUMBER: 14427
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4283
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2040 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYDROTHERMAL: NO
ANTI-SENSE: No

```

Query: Maf012222.1 55.0% Score: 1422.21 E: 2 Length: 2030
Post Index: Similitude: 76.8% Pos: 476-222
Matches: 14226 Complementary: 0 Mismatches: 1631 Indels: 649 Gaps: 127

```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

1  REFERENCE/PROJECT NUMBER: P31 0436
2  TELECOMMUNICATION INFORMATION:
3  TELEPHONE: (619) 535-9001
4  TELEFAX: (619) 535-8949
5  INFORMATION FOR SP2 IN NO: 7:
6  SEQUENCE CHARACTERISTICS:
7  LENGTH: 1659 base pairs
8  TYPE: nucleic acid
9  STRANDEDNESS: single
10 TOPOLOGY: linear
11 MOLECULE TYPE: cDNA
12 IMMEDIATE SOURCE:
13 CLONING: Xp2 (Xp2, SP3)
14 FEATURE:
15 NAME/KEY: CDS
16 LOCATION: 148..1470
17 UN-08-453-694-7

```

Query Match	28.38;	Score 526;	DB 1;	Length 1659;
Best Local Similarity	67.64;	Prod No. 4.8e-115;		
Matches 798;	Conservative 0;	Mismatches 325;	Indels 57;	Gaps 2

[illegible]

DB 1009 ATGGAGCTGATGTTTGTGAAACATCTGGSAGCTAAACGCTGGAGTGCAGATATGACC 1068
UY 1070 TTCTTGAAGGACTTCACTTAAATGAAGAGGACCTTCGACGCGCTTGAGAGTGGAA 1129
DB 1069 TTGCTCAAGGATTTAGTTATACGCGGAGAGCTTTGCCAAGGAGAGGTTGAAAGTGGAA 1128
UY 1130 TTGATCAATCGCATCTTCGAACTTTCGGGGGCGATGGAGAGGCTGGGCTTGAGATGCA 1189
DB 1129 TTCAATACCCGACCTTCGAGTTCGCAAGGCGATGAATGAGCTGCCAGCTCAATGATGCC 1188
UY 1190 GAGTATGCTTTCCTTATCGGATCAACATCTTCGAGCGGATGCGGCTAATGTCAGGAG 1249
DB 1189 GAGTTGGCTTGGTCATTCTATCAGCATCTTCTGTGAGAGCGGCGCAAGTGTAGGAG 1248
UY 1250 CTTAAGCTTGTGAGAGGCTGTATAGGCTTAAGGAGGCGCTGCTGCTTGTACAGGAG 1309
DB 1249 CAGCTGCACTGGAGAGAGCTTCAAGCAATATGCGAAGGCTGCAATGCTATGCTCG 1308
UY 1310 ATCAAGGAGGAGAGAGAGAGAGCTTGGGCTTCGAGGATTCGATGAAGTGGTGAAGCTG 1369
DB 1309 ATGCATCATCTCCATGACCGAATGATGTTCCGACGGATGCTAATGAAT 1368
UY 1370 CGCAGGCTGAGCTCCGCTGAGCTGAGAGAGAGTCTTTCATTCGAGCTGAGAGAGAGAG 1429
DB 1369 CGGAGGCTGAGAGAGGCTGAGCTGAGAGAGAGTGTTCAGCTGGGCTGTGAGAGAGAGAG 1428
UY 1430 CTGCGGCTGCTTGGCTGTGAGAGAGCTGGAGATGTGAGCAGCT 1469
DB 1429 GTCGACGCTGCTGCTGCTGAGAGAGCTGGAGATGTGAGCAGCT 1468

Search completed: April 29, 2001, 16:16:34
Job time: 5497 Sec

